

# L'estimation de modèles avec changements structurels multiples

Pierre Perron

Volume 73, numéro 1-2-3, mars-juin-septembre 1997

L'économétrie appliquée

URI : <https://id.erudit.org/iderudit/602236ar>

DOI : <https://doi.org/10.7202/602236ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)

1710-3991 (numérique)

[Découvrir la revue](#)

Citer cet article

Perron, P. (1997). L'estimation de modèles avec changements structurels multiples. *L'Actualité économique*, 73(1-2-3), 457-505.  
<https://doi.org/10.7202/602236ar>

Résumé de l'article

Cette étude considère le problème de l'estimation de modèles de régressions linéaires avec changements structurels multiples. Nous passons en revue la classe de modèles analysée par Bai et Perron (1996) et certains de leurs résultats asymptotiques. Nous discutons plus en détail un algorithme de calcul, basé sur les principes de la programmation dynamique, qui permet d'obtenir des estimations de façon très efficace même si le nombre de points de rupture est élevé. Ensuite, nous discutons du problème d'estimation de ce nombre de changements via certains critères d'information. Des résultats de simulations sont présentés pour illustrer les mérites et les défauts de ces procédures. Finalement, certains résultats empiriques mettent en évidence l'importance pratique de nos résultats.

## L'ESTIMATION DE MODÈLES AVEC CHANGEMENTS STRUCTURELS MULTIPLES\*

Pierre PERRON

*Centre de recherche et développement en économique (C.R.D.E.)*

*Département de sciences économiques*

*Université de Montréal*

**RÉSUMÉ** – Cette étude considère le problème de l'estimation de modèles de régressions linéaires avec changements structurels multiples. Nous passons en revue la classe de modèles analysée par Bai et Perron (1996) et certains de leurs résultats asymptotiques. Nous discutons plus en détail un algorithme de calcul, basé sur les principes de la programmation dynamique, qui permet d'obtenir des estimations de façon très efficace même si le nombre de points de rupture est élevé. Ensuite, nous discutons du problème d'estimation de ce nombre de changements via certains critères d'information. Des résultats de simulations sont présentés pour illustrer les mérites et les défauts de ces procédures. Finalement, certains résultats empiriques mettent en évidence l'importance pratique de nos résultats.

**ABSTRACT** – This paper considers the problem of estimation in the linear regression model with multiple structural changes. We first survey the class of models analyzed by Bai and Perron (1996) and some of their asymptotic results. We then discuss in greater details a numerical algorithm, based on the principle of dynamic programming, that permits obtaining estimates of the break dates very efficiently even if there is a large number of changes. We also discuss issues related to the estimation of the number of breaks using information criteria. Simulation results are presented to illustrate the merits and drawbacks of such procedures. Finally, some empirical examples highlight the practical importance of our results.

---

\* Je tiens tout d'abord à remercier Jushan Bai. Notre travail conjoint et nos conversations sont la base de cette étude. Merci aussi à Gabriel Rodriguez pour son assistance à la préparation des résultats de simulations et d'estimations et à Cosme Vodounou pour son aide à la préparation du texte. Je souligne également le support financier du Conseil de recherches en sciences humaines du Canada, du Conseil de recherches en sciences naturelles et en génie du Canada et le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec. Ce travail est dédié à la mémoire de Lise Salvas-Bronsard.

## INTRODUCTION

La littérature statistique et économétrique recèle un volume important d'études reliées au problème de changement structurel. Pour les modèles de régressions linéaires, nous rappelons, entre autres, les travaux pionniers de Chow (1960), Quandt (1958) et Barten et Salvat-Bronsard (1970). Depuis ces travaux, qui traitent la date de rupture comme étant connue, l'agenda des recherches s'est dirigé vers une modélisation où cette date de rupture est traitée comme une variable inconnue. Le travail classique de Brown, Durbin et Evans (1975) a généralisé l'approche de type CUSUM aux modèles de régressions linéaires. Bhattacharya (1994), Krishnaiah et Miao (1988) et Deshayes et Picard (1986) présentent un survol des développements en statistique.

Par rapport au problème des tests de changement structurel, les contributions récentes les plus importantes sont les travaux de Andrews (1993) et ceux de Andrews et Ploberger (1994) qui considèrent divers tests ayant certaines propriétés d'optimalité. Dans le cadre de modèles estimés par la méthode de moments généralisée, Sowell (1996) discute d'une classe de tests ayant aussi des propriétés d'optimalité asymptotique. Des études récentes ont aussi abordé ce problème dans le cadre de modèles économétriques avec comme régresseurs des variables avec tendance, racine unitaire, et/ou corrélation temporelle. Des modèles avec cointégration ont également été étudiés. De plus, le problème de changement structurel a reçu beaucoup d'attention en relation avec le débat portant sur la question de racine unitaire versus le changement dans la fonction de tendance pour une série univariée (voir Perron, 1989). Des méthodes de test pour un changement dans la fonction de tendance permettant d'être agnostique par rapport à la présence ou non d'une racine unitaire sont présentées dans Perron (1991) et Vogelsang (1993). Les travaux de Bai (1994, 1995) et Bai, Lumsdaine et Stock (1994) contiennent des résultats importants concernant la distribution asymptotique de l'estimation de la date de rupture lorsqu'un seul changement se produit.

Bien que la littérature soit vaste, les études qui se sont directement intéressées au problème relatif aux changements structurels multiples sont peu nombreuses. À cet effet, nous pouvons citer les travaux de Yao (1988), Yao et Au (1989) et Yin (1988) qui étudient l'estimation du nombre de changements dans la moyenne d'une séquence de variables utilisant le critère d'information de type Bayésien (BIC). Pour des modèles économétriques plus généraux, Garcia et Perron (1994) proposent un test de type Wald pour la présence de deux changements dans une série dynamique. Andrews, Lee et Ploberger (1996) considèrent des tests optimaux dans le modèle linéaire avec variance connue. Liu, Wu et Zidek (1995) considèrent des changements multiples dans le modèle linéaire estimé par moindres carrés et proposent un certain critère d'information pour la sélection du nombre de ruptures. Leurs résultats sont généralisés par Bai et Perron (1996) qui considèrent le problème d'estimation et d'inférence dans un modèle linéaire très général permettant la présence de plusieurs changements structurels. Bai et Perron (1997) discutent plus en détail les problèmes pratiques d'estimations et présentent des résultats de simulations sur les propriétés en petits échantillons d'estimateurs, de tests et de procédures de sélection du nombre de changements.

Cette étude a pour but de présenter un survol des principaux résultats de Bai et Perron (1996, 1997) sur l'estimation des paramètres dans le modèle linéaire avec changements structurels multiples. De plus, nous présentons des résultats de simulations concernant les propriétés des méthodes de sélection du nombre de changements fondées sur divers critères d'information.

Le plan de l'étude est comme suit. Dans la première section, nous présentons le modèle, les postulats et les résultats principaux de convergence de Bai et Perron (1996). Dans la section 2, nous discutons en détail un algorithme, basé sur les principes de la programmation dynamique, lequel algorithme permet d'estimer de façon très efficace des modèles avec plusieurs ruptures. La section 3 présente diverses méthodes de sélection du nombre de changements et fait état des résultats de simulations pour les méthodes basées sur quelques critères d'information. Enfin, les sections 4 et 5 présentent respectivement quelques exemples empiriques et quelques commentaires.

## 1. LE MODÈLE ET LES POSTULATS

Dans cette section, nous présentons le modèle statistique considéré et les postulats imposés sur les variables et les erreurs. Ensuite, nous faisons un bref sommaire des principaux résultats asymptotiques de Bai et Perron (1996).

### 1.1 Le modèle

Nous considérons la régression linéaire multiple suivante avec  $m$  changements (ou  $m + 1$  régimes) :

$$\begin{aligned} y_t &= x'_t \beta + z'_t \delta_1 + u_t, & t = 1, \dots, T_1, \\ y_t &= x'_t \beta + z'_t \delta_2 + u_t, & t = T_1 + 1, \dots, T_2, \\ &\vdots & \\ y_t &= x'_t \beta + z'_t \delta_{m+1} + u_t, & t = T_m + 1, \dots, T. \end{aligned} \tag{1}$$

Dans ce contexte,  $y_t$  est la variable dépendante au temps  $t$ ,  $x_t$  ( $p \times 1$ ) et  $z_t$  ( $q \times 1$ ) sont des vecteurs de régresseurs et  $\beta$  et  $\delta_j$  ( $j = 1, \dots, m + 1$ ) sont les vecteurs de coefficients correspondants ;  $u_t$  est le terme d'erreur au temps  $t$ . Les indices ( $T_1, \dots, T_m$ ) sont les dates de changement et sont traités comme des inconnues de façon explicite. Notons qu'il s'agit d'un modèle à changement structurel partiel étant donné que  $\beta$  n'est pas sujet à changement et est effectivement estimé en utilisant l'échantillon complet. Lorsque  $p = 0$ , nous obtenons un modèle à changement structurel pur où tous les coefficients sont sujets à changement.

Pour l'analyse qui suivra, il est utile d'introduire une certaine terminologie. Nous appelons une  $m$ -partition (ou simplement une partition) des entiers  $(1, \dots, T)$ , un vecteur de dimension  $m$  d'entiers  $(T_1, \dots, T_m)$  tel que  $1 < T_1 < \dots < T_m < T$ .

Pour faciliter l'exposition, nous utiliserons la convention  $T_0 = 0$  et  $T_{m+1} = T$ . Aussi, nous définissons la matrice bloc-diagonale

$$\bar{Z} = \begin{pmatrix} Z_1 & & & \\ & Z_2 & & \\ & & \ddots & \\ & & & Z_{m+1} \end{pmatrix}$$

avec  $Z_i = (z_{T_{i-1}+1}, \dots, z_{T_i})'$ . Nous disons que  $\bar{Z}$  partitionne diagonalement  $Z = (z_1, \dots, z_T)$  à la  $m$ -partition  $(T_1, \dots, T_m)$ . Avec ces définitions, le système de régression linéaire multiple (1) peut s'écrire

$$Y = X\beta + \bar{Z}\delta + U$$

où  $Y = (y_1, \dots, y_T)'$ ,  $X = (x_1, \dots, x_T)'$ ,  $U = (u_1, \dots, u_T)'$ ,  $\delta = (\delta'_1, \delta'_2, \dots, \delta'_{m+1})'$ . Nous notons la vraie valeur d'un paramètre par un 0 en indice supérieur. En particulier,  $\bar{Z}^0$  est la partition diagonale de  $Z$  à la  $m$ -partition  $(T_1^0, \dots, T_m^0)$  correspondant aux vraies dates de rupture. Donc, le processus générateur de données est supposé être

$$Y = X\beta^0 + \bar{Z}^0\delta^0 + U.$$

Le but est d'estimer les inconnues  $(\beta^0, \delta_1^0, \dots, \delta_{m+1}^0, T_1^0, \dots, T_m^0)$  sous l'hypothèse que  $\delta_i^0 \neq \delta_{i+1}^0$  ( $1 \leq i \leq m$ ). En général, le nombre de points de rupture  $m$  peut être considéré comme inconnu avec  $m^0$  pour vraie valeur. Cependant, pour l'instant, nous le traitons comme connu.

## 1.2 La méthode d'estimation

La méthode d'estimation considérée est basée sur le principe de la minimisation des erreurs au carré. Considérons tout d'abord une quelconque  $m$ -partition  $(T_1, \dots, T_m)$ . Pour cette  $m$ -partition, les estimations MCO de  $\beta$  et  $\delta_j$  sont obtenues en minimisant la somme des résidus au carré

$$(Y - X\beta - \bar{Z}\delta)' (Y - X\beta - \bar{Z}\delta) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} [y_t - x_t'\beta - z_t'\delta_i]^2$$

Dénotons par  $\hat{\beta}(\{T_j\})$  et  $\hat{\delta}(\{T_j\})$  les estimations obtenues en utilisant la  $m$ -partition  $(T_1, \dots, T_m)$  dénotée ici par  $\{T_j\}$ . La somme des carrés des erreurs correspondant à ces estimations est donc

$$S_T(T_1, \dots, T_m) = (Y - X\hat{\beta}(\{T_j\}) - \bar{Z}\hat{\delta}(\{T_j\}))' (Y - X\hat{\beta}(\{T_j\}) - \bar{Z}\hat{\delta}(\{T_j\})).$$

Les estimations des points de rupture  $(T_1^0, \dots, T_m^0)$  sont les dates  $(\hat{T}_1, \dots, \hat{T}_m)$  qui permettent d'obtenir un minimum global de  $S_T(T_1, \dots, T_m)$  par rapport à toutes les  $m$ -partitions possibles  $(T_1, \dots, T_m)$  telles que  $T_i - T_{i-1} \geq q$ . C'est-à-dire

$$(\hat{T}_1, \dots, \hat{T}_m) = \arg \min_{T_1, \dots, T_m} S_T(T_1, \dots, T_m). \quad (2)$$

Les estimations des dates de rupture sont donc les minimisations globales de la fonction objectif. Finalement, les estimations des coefficients des régressions sont simplement les estimations MCO obtenues en utilisant la  $m$ -partition  $\{\hat{T}_j\}$ , c'est-à-dire

$$\hat{\beta} = \hat{\beta}(\{\hat{T}_j\}), \hat{\delta} = \hat{\delta}(\{\hat{T}_j\}). \quad (3)$$

Puisque les dates de changement sont des variables discrètes, il est possible d'obtenir les minimiseurs globaux, la solution du système (3), par un balayage exhaustif par rapport à toutes les partitions possibles. Cette méthode devient rapidement excessive en temps de calcul lorsque  $m > 2$ , puisque le nombre possible de  $m$ -partition est d'ordre  $T^m$ . Heureusement, il existe une méthode très efficace, basée sur les principes de la programmation dynamique, qui permet d'obtenir des minimiseurs globaux en effectuant tout au plus des opérations MCO d'ordre  $T^2$ , quel que soit  $m$ . Cette méthode est discutée en détail à la section 2.

### 1.3 Les postulats

Pour que nos estimations aient des propriétés asymptotiques intéressantes (convergence, etc.), nous devons imposer certaines contraintes aux régresseurs et aux erreurs. Certains postulats sont de nature purement technique et n'imposent pas de contraintes en pratique. D'autres, cependant, limitent l'applicabilité des résultats à une certaine classe de modèles. Pour faciliter l'exposition, nous utilisons la notation  $W_i = (x'_i, z'_i)$ ,  $W = (w_1, \dots, w_T)'$  et  $\bar{W}^0$  la partition diagonale de  $W$  à la  $m$ -partition  $(T_1^0, \dots, T_m^0)$  telle que  $\bar{W}^0 = \text{diag}(W_1^0, \dots, W_{m+1}^0)$ . Les trois premiers postulats sont de nature purement technique et ne sont pas contraignants en pratique.

**A1**  $T_i^0 = [T\lambda_i^0]$  avec  $0 \equiv \lambda_0 < \lambda_1 < \dots < \lambda_m^0 < \lambda_{m+1}^0 \equiv 1$  où  $[\cdot]$  dénote la partie entière.

**A2** Il existe un  $\ell_0$  tel que pour tout  $\ell > \ell_0$  les valeurs propres minimales des matrices  $\frac{1}{\ell} \sum_{T_i^0+1}^{T_i^0+\ell} w_i w_i'$  et  $\frac{1}{\ell} \sum_{T_i^0-\ell}^{T_i^0} w_i w_i'$  sont bornés au-dessus de zéro ( $i = 1, \dots, m+1$ ).

**A3** La matrice  $A_{k\ell} = \sum_k^\ell z_i z_i'$  est inversible pour  $\ell - k \geq q$ , la dimension du vecteur  $z_i$ .

Le postulat A1 spécifie la nature du cadre asymptotique adopté pour discuter de concepts tels que la convergence. Il spécifie que les points de rupture sont asymptotiquement distincts et que chaque segment (ou régime) croît en proportion avec la taille de l'échantillon total. Les quantités  $\lambda^0 = (\lambda_1^0, \dots, \lambda_m^0)$  sont appelées

les « fractions de rupture ». Le postulat A2 spécifie qu'il y a assez d'observations autour des points de rupture pour que ces derniers puissent être identifiés. Le postulat A3 permet d'appliquer la méthode de MCO pour toutes les  $m$ -partitions  $(T_1, \dots, T_m)$  telles que  $T_i - T_{i-1} \geq q$ . Notons que ce postulat peut être évité en utilisant une méthode d'estimation basée sur les inverses généralisées.

Le prochain postulat est plus contraignant quant à la nature des variables introduites dans le modèle.

**A4** Pour tout  $i = 1, \dots, m + 1$ ,  $W_i^0{}'W_i^0/(T_i^0 - T_{i-1}^0)$  converge en probabilité vers une matrice positive définie non stochastique (pas nécessairement la même pour tous les  $i$ ).

Ce postulat est standard pour le modèle classique de régression linéaire mais contraignant, puisqu'il spécifie qu'on ne peut avoir de variables intégrées (avec une racine unitaire autorégressive). Il permet cependant des variables ayant une tendance déterministe où l'indice de temps est dégonflé par  $T$ . Par exemple, pour une tendance linéaire  $g(t) = a + b(t/T)$ .

Le dernier postulat concerne la nature des erreurs  $u_t$  en relation avec celle des régresseurs  $\{x_t, z_t\}$ . Deux cas sont permis :

**A5(i)** Soit  $\{\mathcal{F}_i, i = 1, 2, \dots\}$  une séquence de  $\sigma$ -algèbres, la paire  $\{u_t, \mathcal{F}_i\}$  forme une séquence de «  $L^r$ -mixingale » avec  $r = 4 + \delta$  pour un certain  $\delta > 0$ . De plus, les erreurs  $u_t$  sont indépendantes des régresseurs  $w_s$  pour tout  $t$  et  $s$ .

**A5(ii)** Soit  $\mathcal{F}_t$ , la  $\sigma$ -algèbre générée par  $\{\dots, w_{t-1}, w_t, \dots, u_{t-2}, u_{t-1}\}$ ; a)  $\{u_t\}$  est une séquence de différences de martingales relativement à  $\{\mathcal{F}_t\}$  et  $\sup_t E|u_t|^{4+\delta} < \infty$ ;

b)  $T^{-1} \sum_{t=1}^{[Tv]} z_t z_t' \rightarrow Q(v)$  uniformément en  $v \in [0, 1]$  avec  $Q(v)$  une matrice définie positive pour  $v > 0$  et strictement croissante en  $v$ ; c) si les erreurs  $u_t$  ne sont pas indépendantes des régresseurs  $\{\beta_s\}$  pour tout  $t$  et  $s$ , le problème de minimisation défini par (2) est contraint à toutes les partitions telles que  $T_i - T_{i-1} > \varepsilon T$  ( $i = 1, \dots, m + 1$ ) pour un  $\varepsilon > 0$ .

Quoique ce postulat impose certaines contraintes, il permet une classe de modèles très générale. Les conditions décrites dans la partie (i) s'adressent au cas où il n'y a pas de variables dépendantes retardées dans les régresseurs  $w_t$ . Dans ce cas, les conditions sur les résidus permettent une classe très générale de corrélations temporelles et d'hétéroscédasticité. Les conditions de la partie (ii) s'adressent au cas où des variables dépendantes retardées sont permises comme régresseurs. Dans ce cas, il ne peut y avoir de corrélations temporelles dans les erreurs, bien qu'un certain degré d'hétéroscédasticité soit permis. Cette généralisation est obtenue au prix d'une certaine restriction additionnelle quant aux  $m$ -partitions permises, ce qui, d'un point de vue pratique, n'est pas contraignant, puisque  $\varepsilon$  peut être arbitrairement petit.

Le choix entre les postulats A5(i) et A5(ii) peut être très intéressant dans le cas des modèles dynamiques lorsque les paramètres associés aux variables dépendantes retardées ne sont pas sujets à changement. Dans ce cas, l'investigateur peut modéliser cette dynamique par une approche paramétrique (en introduisant assez de variables dépendantes retardées pour avoir des résidus non corrélés) ou par une approche indirecte non paramétrique (en laissant les effets de dynamique dans les résidus et en appliquant une correction non paramétrique pour conduire à une inférence asymptotique appropriée). Ce choix de modélisation peut être fort utile selon que le changement se produit rapidement ou de façon graduelle. Prenons, par exemple, le cas d'un changement de moyenne dans une série temporelle corrélée. En spécifiant  $z_t = \{1\}$ ,  $x_t = \{\emptyset\}$ , toute la dynamique est contenue dans le terme d'erreurs et n'affecte pas l'effet d'un changement de moyenne sur le niveau de la série (le changement est donc abrupt). Cependant, en spécifiant  $z_t = \{1\}$  et  $x_t = \{\text{retards de } y_t\}$ , un changement dans le coefficient associé à la constante  $z_t$  est relié à un changement dans le niveau de  $y_t$  qui varie pour les périodes suivant la date de rupture. Cette variation dépend de la dynamique autorégressive et s'effectue de façon graduelle vers un niveau limite.

#### 1.4 Les résultats asymptotiques

Bai et Perron (1996) ont dérivé certains résultats asymptotiques concernant les estimations  $\hat{\beta}$ ,  $(\hat{\delta}_1, \dots, \hat{\delta}_{m+1})$  et  $(\hat{T}_1, \dots, \hat{T}_m)$  sous les conditions décrites à la section 1.3. Nous en faisons ici un bref sommaire.

Les résultats les plus importants concernent la convergence des quantités  $\hat{\lambda}_i = \hat{T}_i / T$  et le rythme de convergence. Le résultat obtenu est que non seulement  $\hat{\lambda}_i$  converge vers sa vraie valeur  $\lambda_i^0$ , mais qu'il le fait au rythme rapide  $T$ , c'est-à-dire  $T(\hat{\lambda}_i - \lambda_i^0) = O_p(1)$  pour tout  $i$ . Il est important, cependant, de remarquer que cette convergence au rythme  $T$  s'applique aux quantités  $\hat{\lambda}_i$  et non aux dates de rupture  $\hat{T}_i$  elles-mêmes. Pour ces derniers, le résultat stipule qu'avec une probabilité arbitrairement proche de 1, la distance entre  $\hat{T}_i$  et  $T_i^0$  est bornée par une certaine constante indépendante de la taille de l'échantillon. Ce phénomène est essentiellement dû au cadre d'analyse asymptotique adopté tel que spécifié par le postulat A1.

Malgré que les dates de rupture  $\hat{T}_i$  ne soient pas convergentes pour les vraies valeurs  $T_i^0$ , le fait que les quantités  $\hat{\lambda}_i$  convergent au rythme rapide d'ordre  $T$  est suffisant pour que l'estimation de ces points de rupture n'affecte pas, en grands échantillons, la distribution limite des estimations des autres paramètres du modèle. Ceci nous permet d'obtenir, pour ces estimations, les résultats classiques de normalité asymptotique au rythme  $\sqrt{T}$ . De façon plus précise, nous avons



$$\begin{pmatrix} \sqrt{T}(\hat{\beta} - \beta^0) \\ \sqrt{T}(\hat{\delta} - \delta^0) \end{pmatrix} \xrightarrow{d} N(0, V^{-1}\Phi V^{-1})$$

où

$$V = p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{pmatrix} X'X & X'\bar{Z}^0 \\ \bar{Z}^{0'}X & \bar{Z}^{0'}\bar{Z}^0 \end{pmatrix}$$

$$\Phi = p \lim_{T \rightarrow \infty} \frac{1}{T} (X, \bar{Z}^0)' \Omega (X, \bar{Z}^0)$$

et

$$\Omega = E(UU').$$

Notons que si les erreurs sont non corrélées et homoscédastiques alors  $\Phi = \sigma^2 V$ , et la matrice de covariance asymptotique se réduit à  $\sigma^2 V^{-1}$ , laquelle peut être estimée de façon convergente en utilisant une estimation convergente de  $\sigma^2$  tel que  $T^{-1} \sum_{t=1}^T \hat{u}_t^2$ . Lorsqu'il y a corrélation et/ou hétéroscedasticité dans les résidus, une estimation convergente de  $\Phi$  peut être construite selon les méthodes proposées par Newey et West (1987) et Andrews (1991). Notons, finalement, que cette correction pour la présence de corrélation peut être faite en supposant des distributions identiques à travers les différents régimes ou en permettant aux distributions des régresseurs et des erreurs de changer.

Le papier de Bai et Perron (1996) présente des résultats par rapport à plusieurs autres sujets : a) la distribution limite des dates de rupture permettant de construire des intervalles de confiance asymptotiques ; b) des statistiques de tests pour les hypothèses suivantes,  $m^0 = 0$  versus  $m^0 = k$  un nombre fixe choisi *a priori*;  $m^0 = 0$  versus un certain nombre inconnu entre 1 et 5 ; et  $m^0 = \ell$  versus  $m^0 = \ell + 1$ . Ce dernier cas est la base d'une méthode séquentielle pour estimer le nombre de points de rupture. Puisque l'objet de la présente étude est l'estimation, nous passons sous silence ces divers résultats.

## 2. MÉTHODE DE CALCUL DES MINIMISEURS GLOBAUX

Dans cette section, nous discutons un algorithme basé sur les principes de la programmation dynamique qui permet d'obtenir efficacement les estimations des points de rupture en tant que minimiseurs globaux de la somme des résidus au carré des erreurs. Cette méthode permet le calcul de ces estimations en effectuant des opérations MCO tout au plus d'ordre  $T^2$ , quel que soit le nombre de changements  $m$ . La méthode n'est pas vraiment nouvelle et a des antécédents

dans la littérature statistique, notamment Guthery (1974), Bellman et Roth (1969) et Fisher (1958). Elle semble cependant avoir été oubliée, à tout le moins dans la littérature économétrique, et une description détaillée paraît donc s'imposer.

### 2.1 La matrice triangulaire de sommes des résidus au carré

L'idée de base de cette approche devient très intuitive lorsqu'on réalise qu'avec un échantillon de taille  $T$ , il ne peut exister plus de  $T(T+1)/2$  segments, quel que soit le nombre de points de rupture  $m$ . Ceci est représenté à la figure 1 où l'axe vertical représente la date initiale d'un segment et l'axe horizontal la date terminale. Chaque entrée représente la somme des carrés des résidus estimés correspondant à ce segment. La somme globale des résidus estimés au carré pour n'importe quelle  $m$ -partition ( $T_1, \dots, T_m$ ) et pour toute valeur de  $m$  doit nécessairement être une certaine combinaison linéaire de ces  $T(T+1)/2$  sommes de résidus au carré. Les estimations des dates de rupture, la  $m$ -partition ( $\hat{T}_1, \dots, \hat{T}_m$ ), correspondent à cette combinaison qui permet d'obtenir une valeur minimale. L'algorithme de programmation dynamique peut être vue comme une méthode efficace pour comparer ces combinaisons de sommes de résidus au carré (correspondant aux différentes  $m$ -partitions) pour obtenir un minimum global.

Avant de discuter de cet algorithme, notons qu'en pratique, moins de  $T(T+1)/2$  segments sont permis. Une distance minimale, soit  $h$  périodes, peut être imposée entre chaque point de rupture. C'est-à-dire que l'utilisateur peut spécifier que chaque segment soit au moins de  $h$  périodes. Notons que  $h < q$  (le nombre de coefficients qui varient) est possible ; dans de tels cas, la somme des résidus au carré des segments de longueur inférieure à  $q$  est identiquement 0. Pour simplifier, on supposera cependant que  $h \geq q$ . Souvent, l'utilisateur peut spécifier cette longueur minimale sans référence à  $q$ , par exemple pour minimiser l'effet possible d'observations aberrantes.

Cette longueur minimale pour chaque segment implique une réduction dans le nombre total de segments à considérer de  $(h-1)T - (h-2)(h-1)/2$  (voir figure 1).

D'autres réductions dans le nombre total de segments à considérer sont possibles. En effet, le segment le plus long doit être tout de même assez court pour permettre la présence de  $m$  autres segments avant et après. Par exemple, si un segment débute à une date entre 1 et  $h$ , la date terminale maximale de ce segment est  $T - hm$  si  $m$  ruptures sont permises (ou  $m+1$  régimes). Ceci permet une réduction additionnelle de  $h^2m(m+1)/2$  segments à considérer.

FIGURE 1

EXEMPLE DE LA MATRICE TRIANGULAIRE DE SOMMES DE RÉSIDUS AUX CARRÉS AVEC  
 $T = 25$ ,  $h = 5$  ET  $m = 2$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
2		$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
3			$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
4				$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
5					$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
6						$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
7							$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
8								$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
9									$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
10										$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$x^b$	$x^b$	$x^b$	$x^b$	$x^b$
11											$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
12												$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
13													$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
14														$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
15															$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
16																$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
17																	$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
18																		$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
19																			$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$	$\bullet$
20																				$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$	$\bullet$
21																					$x^a$	$x^a$	$x^a$	$x^a$	$\bullet$
22																						$x^a$	$x^a$	$x^a$	$x^a$
23																							$x^a$	$x^a$	$x^a$
24																								$x^a$	$x^a$
25																									$x^a$

NOTES :

$x^a$  indique les segments rejetés puisqu'un segment doit être au moins de longueur 5,

$x^b$  indique les segments rejetés puisqu'autrement il n'y aurait pas de place pour 3 segments de longueur 5.

Un • indique un segment permissible.

En somme, toute l'information pertinente au calcul du minimum global de la somme des résidus au carré peut être obtenue en examinant les sommes des résidus au carré correspondant à

$$T(T+1)/2 - (h-1)T + (h-2)(h-1)/2 - h^2m(m+1)/2$$

segments. Nous devons calculer la somme des résidus au carré des segments ayant les dates initiales et terminales suivantes :

date initiale

date terminale

$$i = \ell h + 1, \dots, (\ell + 1)h \quad j = h + i - 1, \dots, T - (m - \ell)h \quad (\ell = 0, \dots, m - 1)$$

$$i = hm + 1, \dots, T - h + 1 \quad j = h + i - 1, \dots, T.$$

## 2.2 Le cas d'un modèle à changement structurel pur

Considérons, en premier lieu, le cas où nous avons un modèle à changement structurel pur, c'est-à-dire  $p = 0$  et le système est décrit par

$$Y = \bar{Z}\delta + U \quad (4)$$

où nous rappelons que  $\bar{Z}$  est une matrice bloc-diagonale, les différents blocs correspondant aux régimes spécifiés par la  $m$ -partition  $(T_1, \dots, T_m)$ . Dans ce cas, le calcul des estimations  $\hat{\delta}$ ,  $\hat{u}_t$  et  $S_T(T_1, \dots, T_m)$  peut être effectué en appliquant la méthode MCO segment par segment sans contrainte à travers ceux-ci.

Dans ce cas, le calcul de la matrice triangulaire de sommes de résidus au carré peut se faire en utilisant des formules standard pour le calcul de résidus récursifs. En fait, toute l'information pertinente peut être calculée à partir de  $T - hm + 1$  ensembles de résidus récursifs. Soit  $v(i, j)$  le résidu récursif au temps  $j$  obtenu en utilisant un échantillon commençant à la date  $i$  et  $SSR(i, j)$  la somme des résidus au carré d'un segment commençant à la date  $i$  et se terminant à la date  $j$ . Nous avons la formule de récurrence suivante (voir Brown, Durbin et Evans, 1975) :

$$SSR(i, j) = SSR(i, j - 1) + v(i, j)^2.$$

Toute l'information pertinente est contenue dans les valeurs de  $SSR(i, j)$  pour les combinaisons  $(i, j)$  indiquées ci-dessus. Notons que tout ce calcul n'implique qu'un nombre d'inversions matricielles d'ordre  $T$ .

## 2.3 L'algorithme de programmation dynamique

Lorsque les sommes des résidus au carré des segments pertinents ont été calculées et mises en mémoire, une approche de programmation dynamique peut être utilisée pour évaluer quelle partition permet d'obtenir une minimisation globale de la somme des résidus au carré sur l'ensemble de l'échantillon. Essentiellement, cette méthode agit via un examen séquentiel de partitions optimales avec un point de rupture (ou deux segments). Soit  $SSR(\{\hat{T}_{r,n}\})$  la somme des résidus au carré associée à la partition optimale contenant  $r$  points de rupture et qui utilise les premières  $n$  observations. La  $m$ -partition optimale pour l'échantillon total peut être obtenue en résolvant le problème de récursion suivant

$$SSR(\{T_{m,T}\}) = \min_{mh \leq j \leq T-h} [SSR(\{T_{m-1,j}\}) + SSR(j+1, T)]. \quad (5)$$

Il est utile d'écrire (5) sous la forme suivante :

$$\begin{aligned}
 SSR(\{T_{m,T}\}) = & \min_{mh \leq j_1 \leq T-h} [SSR(j_1 + 1, T) + \\
 & \min_{(m-1)h \leq j_2 \leq j_1-h} [SSR(j_2 + 1, j_1) + \\
 & \min_{(m-2)h \leq j_3 \leq j_2-h} [SSR(j_3 + 1, j_2) + \\
 & \vdots \\
 & \min_{h \leq j_m \leq j_{m-1}-h} [SSR(1, j_m) + SSR(j_m + 1, j_{m-1})] \dots ]].
 \end{aligned}$$

En examinant ce problème de minimisation, on remarque que la procédure commence par évaluer la 1-partition optimale pour tous les sous-échantillons qui permettent une rupture s'étalant de l'observation  $h$  à l'observation  $T - mh$ . Donc, dans un premier temps, il faut sauvegarder un ensemble de  $(T - (m + 1)h + 1)$  1-partitions optimales avec les sommes des résidus au carré qui leur sont associées. Chacune des 1-partitions optimales correspond à des sous-échantillons avec date finale s'échelonnant de  $2h$  à  $T - (m - 1)h$ .

Considérons maintenant la prochaine étape qui consiste à rechercher des partitions optimales avec deux ruptures (2-partitions). De telles partitions ont des dates terminales s'échelonnant de l'observation  $3h$  à l'observation  $T - (m - 2)h$ . Pour chaque date terminale possible, la procédure recherche laquelle des 1-partitions optimales (sauvegardées à l'étape précédente) peut être insérée pour aboutir à une somme des résidus au carré minimale. Le résultat est un ensemble de  $(T - (m + 1)h + 1)$  2-partitions optimales (ou des partitions optimales avec 3 segments). La méthode continue de façon séquentielle jusqu'à ce qu'un ensemble de  $(T - (m + 1)h + 1)$   $(m - 1)$ -partitions optimales soient obtenues, correspondant à des dates terminales s'échelonnant de l'observation  $(m - 1)h$  à l'observation  $T - 2h$ . L'étape finale est de voir laquelle de ces  $(m - 1)$ -partitions optimales permet d'obtenir une somme globale des résidus au carré lorsque combinée avec un segment additionnel.

On peut donc décrire la méthode comme une mise à jour séquentielle de  $T - (m + 1)h + 1$  segments en partitions optimales avec un, deux et jusqu'à  $(m - 1)$  points de rupture (ou avec deux, trois et jusqu'à  $m$  sous-segments), la dernière étape étant simplement de créer une seule partition optimale avec  $m$  points de rupture (ou  $(m + 1)$  segments), soit une  $m$ -partition optimale.

Il est important de noter qu'en pratique, cette méthode est très rapide avec des échantillons de tailles usuelles. En fait, la composante majeure du temps de calcul est la construction de la matrice triangulaire de sommes des résidus au carré pour tous les segments possibles. La recherche de la  $m$ -partition optimale représente une addition en temps de calcul qui est comparativement marginale. Ceci veut dire qu'il est à peine plus long d'obtenir des minimiseurs globaux avec 5 ou 10 points de rupture qu'il ne l'est avec 2 points de rupture.

## 2.4 Le cas de modèles à changement structurel partiel

La méthode décrite ci-dessus ne peut s'appliquer directement au cas d'un modèle à changement structurel partiel, c'est-à-dire avec  $p > 0$ . Ceci est dû au fait que nous ne pouvons estimer le vecteur de paramètre  $\beta$  sans connaître la  $m$ -partition optimale. Autrement dit, l'estimation de  $\beta$  associée à la minimisation globale dépend elle-même de la  $m$ -partition globale que l'on recherche. Contrairement au cas d'un modèle à changement structurel pur, lorsqu'on peut écrire la régression sous la forme (4), chaque élément de la matrice triangulaire de sommes de résidus au carré dépend de la  $m$ -partition optimale finale que l'on recherche. Il existe cependant une méthode itérative simple qui permet de pallier ce problème.

La procédure récursive peut être décrite de la façon suivante. En posant  $\theta = (\delta, T_1, \dots, T_m)$ , nous pouvons écrire la somme des résidus au carré en fonction de  $\beta$  et  $\theta$ , c'est-à-dire  $SSR(\beta, \theta)$ . Adoptant la méthode de Sargan (1964), nous pouvons minimiser  $SSR(\beta, \theta)$  comme suit : tout d'abord, minimiser par rapport à  $\theta$  en gardant  $\beta$  fixe, ensuite minimiser par rapport à  $\beta$  en gardant  $\theta$  fixe, et enfin itérer la procédure. Chaque itération assure une baisse de la fonction objectif. Les propriétés de convergence de cette méthode itérative sont discutées dans Sargan (1964). Bien entendu, la convergence vers le minimum global n'est pas assurée et, dans ce contexte, le choix de la valeur initiale de  $\beta$  peut être important pour éviter un minimum local.

Nous discutons les détails de cette méthode dans notre contexte et nous effectuons aussi une légère modification qui permet une convergence très rapide. Nous remarquons que la première étape relative à la minimisation par rapport à  $\theta$  en gardant  $\beta$  fixe, équivaut simplement à l'application de l'algorithme de programmation dynamique discutée ci-dessus, mais, cette fois-ci, avec  $y_t - x_t'\beta$  comme variable dépendante. Puisque  $\beta$  est fixe, cette étape équivaut en fait à un modèle à changement structurel pur. Soit  $\theta^* = (\delta^*, \{T^*\})$  les estimations obtenues de cette première étape [avec  $\{T^*\} = (T_1^*, \dots, T_m^*)$ ]. L'application de la méthode de Sargan suggère que la deuxième étape soit une simple régression linéaire avec variable dépendante  $y_t - z_t'\delta_j^*$  pour  $t$  dans le régime  $j$  ( $j = 1, \dots, m + 1$ ) (les régimes ou segments étant définis par la  $m$ -partition  $\{T^*\}$ ) et avec variables dépendantes  $x_t$ .

Des gains énormes en efficacité peuvent être obtenus en apportant une modification mineure à cette seconde étape. Il s'agit de garder seulement  $\{T^*\}$  fixe et de maximiser à nouveau par rapport à  $\delta$  (en plus de  $\beta$ ). Donc,  $\delta$  est mise à jour à chacune des deux étapes. La raison qui justifie ce gain d'efficacité peut être expliquée comme suit : en général, les valeurs de  $\{T^*\}$  obtenues à la première itération seront très près des valeurs  $\{\hat{T}\}$  correspondant au minimum global (à moins que la valeur initiale de  $\beta$  soit très éloignée de la vraie valeur  $\beta^0$ ). Intuitivement, une mauvaise spécification dans la valeur initiale de  $\beta$  n'a que peu d'effet sur les estimations  $\{T^*\}$ , puisque ceux-ci dépendent surtout des changements du vecteur  $\delta$  à travers les différents régimes qui eux sont associés aux

variables  $z_i$ . Considérons une deuxième étape qui applique une régression linéaire ordinaire de la forme

$$Y = X\beta + \bar{Z}^* \delta + U,$$

avec  $\bar{Z}^*$  la partition diagonale de  $Z$  à la  $m$ -partition  $(T_1^*, \dots, T_m^*)$ . Si les valeurs de  $\{T^*\}$  sont égales aux valeurs  $\{\hat{T}\}$  correspondant au minimum global, les estimations de  $\beta$  et de  $\delta$  de cette deuxième étape seront immédiatement celles correspondant au minimum global. Des essais pour tester cette procédure ont démontré que dans la majorité des cas, une seule itération suffit. Dans quelques cas, deux itérations étaient nécessaires, mais il était difficile de trouver des exemples où plus de 2 itérations étaient requises.

Pour faire ressortir le contraste entre ces deux approches, considérons ce qui se produit si  $\delta$  n'est pas mis à jour à nouveau à la deuxième étape. Cette deuxième étape devient donc une régression linéaire simple de la forme

$$Y - \bar{Z}^* \delta^* = X\beta + U.$$

Notons que même si  $\{T^*\}$  est égale à  $\{\hat{T}\}$ , correspondant au minimum global, l'estimation de  $\beta$  ne correspondra pas à  $\hat{\beta}$  (associé au minimum global), à moins, bien sûr, que  $\delta^*$  ait déjà atteint  $\hat{\delta}$  (ce qui ne peut se produire qu'avec une très faible probabilité). Le besoin de faire des itérations additionnelles s'est donc imposé et les expériences menées ont démontré que le nombre d'itérations nécessaires pour arriver aux valeurs correspondant au minimum global étaient très élevé, même dans les modèles très simples.

Le critère de convergence adopté est que le changement dans la valeur objectif  $SSR(T_1, \dots, T_m)$  soit plus petit qu'un certain  $\epsilon$ . En adoptant la méthode itérative suggérée, il est possible de spécifier  $\epsilon = 0$  à cause de la nature discrète des variables  $(T_1, \dots, T_m)$ . En fait, dans la plupart des simulations que nous avons faites, le minimum était atteint après la première itération et la seconde ne faisait que vérifier qu'il n'y avait effectivement aucun changement dans la fonction objectif.

L'efficacité de la méthode proposée ci-dessus pour atteindre le minimum global (par opposition à un certain minimum local) repose sur un choix approprié de la valeur initiale du vecteur  $\beta$ . Nous suggérons la procédure suivante qui s'avère être très efficace. Il s'agit tout d'abord de considérer tous les coefficients comme étant sujets à des changements structurels, même ceux du vecteur  $\beta$ . Donc, nous considérons le modèle comme en étant un à changement structurel pur auquel nous pouvons directement appliquer l'algorithme de programmation dynamique pour obtenir les estimations. Pour être plus précis, écrivons ce modèle à changement structurel pur sous la forme

$$y_t = (x'_t, z'_t)(\delta_{1j}, \delta_{2j})' + u_t \quad t = T_{j-1} + 1, \dots, T_j,$$

pour  $j = 1, \dots, m + 1$ . L'application de l'algorithme de programmation dynamique à ce modèle donne les estimations  $(\delta_{1j}^a, \delta_{2j}^a; j = 1, \dots, m + 1)$  et  $(T_1^a, \dots, T_m^a)$ . Pour

obtenir une valeur initiale du vecteur  $\beta$ , il s'agit simplement d'utiliser la régression MCO suivante :

$$Y - \bar{Z}^a \delta = X\beta + U,$$

où  $\bar{Z}^a$  est la partition diagonale de  $Z$  à la  $m$ -partition  $(T_1^a, \dots, T_m^a)$ . L'estimation de  $\beta$  obtenue ainsi, disons  $\beta^a$ , sert de valeur initiale pour commencer la récursion décrite ci-dessus.

L'adoption de cette méthode pour choisir la valeur initiale de  $\beta$  est justifiée par le fait que les estimations  $\lambda_j^a$  des proportions  $\lambda_j^0$  sont convergentes même lorsque certains coefficients ne changent pas d'un régime à l'autre. La propriété de convergence est assurée dès qu'il y a un coefficient qui change à chaque point de rupture. Donc, la valeur initiale  $\beta^a$  ainsi obtenue est, d'un point de vue asymptotique, équivalente à l'estimation associée à la minimisation globale. Ceci permet d'atteindre le minimum de la fonction objectif en très peu d'itérations et cette approche minimise grandement le risque de convergence vers un minimum local. En fait, ce problème ne s'est produit dans aucun des cas que nous avons analysés (par simulations ou avec des données).

Évidemment, il se peut que cette procédure proposée pour choisir la valeur initiale de  $\beta$  soit difficile ou même impossible à mettre en oeuvre dans certains cas. Nous pensons ici à certains cas où les dimensions  $p$  du vecteur  $\beta$  et  $m$  du nombre de ruptures sont grandes. Dans ces situations, il reste toujours le choix d'utiliser des valeurs initiales fixes pour  $\beta$ . Ici, le problème de convergence vers un minimum local devient, comme toujours, présent et une analyse de sensibilité s'avère de rigueur.

### 3. SÉLECTION DU NOMBRE DE RUPTURES PAR DES CRITÈRES D'INFORMATION

Un problème important pour l'analyse des modèles avec changements structurels multiples est l'estimation du nombre de changements présents. Deux classes générales de procédures ont été proposées. Celle suggérée par Bai et Perron (1996) consiste en une méthode séquentielle de tests. Elle consiste en l'application de tests pour  $\ell$  versus  $\ell + 1$  changements pour une séquence avec  $\ell = 0, 1, 2, \dots$ . L'estimation du nombre de changements est la première valeur de  $\ell$  qui correspond à un non-rejet de l'hypothèse nulle.

L'autre classe de procédures, qui nous intéresse plus particulièrement ici, consiste à utiliser un critère d'information comme méthode de sélection de modèle. L'idée de base est que l'ajout de points de rupture diminue nécessairement la somme des résidus au carré de la régression. Il faut donc pénaliser, d'une quelconque façon, l'ajout de chaque point de rupture additionnel. Les critères de sélection proposés dans la littérature diffèrent essentiellement par le facteur de pénalité. Dans le cadre de modèles avec changements structurels dans la moyenne, Yao (1988) suggère l'utilisation du critère d'information de type Bayésien (*BIC*) défini par



$$BIC(m) = \ell n(\hat{\sigma}^2(m)) + p^* \ell n(T)/T,$$

où  $\hat{\sigma}^2(m) = T^{-1}S_T(\hat{T}_1, \dots, \hat{T}_m)$  et  $p^* = (m+1)q + m + p$ . Le nombre de changements sélectionné  $\hat{m}_{BIC}$  est tel que

$$\hat{m}_{BIC} = \arg \min_m BIC(m).$$

Yao (1988) a démontré que  $\hat{m}_{BIC}$  est convergent pour la vraie valeur  $m^0$ . Un critère souvent utilisé en économétrie est celui d'Akaike défini par

$$AIC(m) = \ell n(\hat{\sigma}^2(m)) + 2p^*.$$

Nous avons répété les expériences de simulations discutées ci-dessus pour le critère  $AIC$ . Cependant, sa performance est si désastreuse qu'il est inutile d'en présenter les résultats.

Plus récemment, Liu, Wu et Zidek (1995) ont proposé une modification du critère de Schwarz de la forme

$$LWZ(m) = \ell n(S_T(\hat{T}_1, \dots, \hat{T}_m)/(T-p^*)) + (p^*/T)c_0(\ell n(T))^{2+\delta_0},$$

avec estimation correspondante

$$\hat{m}_{LWZ} = \arg \min_m LWZ(m).$$

Se basant sur la performance de  $\hat{m}_{LWZ}$  lors de certaines expériences de simulations avec divers modèles, ils suggèrent d'utiliser les valeurs  $\delta_0 = 0,1$  et  $c_0 = 0,299$ .

Le but des simulations présentées dans cette section est d'évaluer la performance des critères d'information  $BIC$  et  $LWZ$  dans le contexte spécifique de détection de rupture dans une fonction de tendance linéaire lorsque les données sont corrélées. La motivation de cet exercice vient de nos travaux antérieurs sur la question de racine unitaire versus changements structurels dans la fonction de tendance pour les données macroéconomiques (voir Perron, 1989, 1994).

Le cadre structurel de base est la décomposition standard d'une série entre fonction de tendance déterministe et partie stochastique (si on fait abstraction de composantes saisonnières), c'est-à-dire

$$y_t = \tau' d_t + v_t,$$

où habituellement  $d_t = \{1\}$  (moyenne non nulle sans croissance) et  $d_t = \{1, t\}$ . Étant donné la nature de la plupart des séries macroéconomiques, il est important de permettre la corrélation temporelle dans la partie stochastique  $v_t$  et, en particulier, la possibilité d'une racine autorégressive unitaire. Pour ce faire, nous adoptons un simple modèle  $AR(1)$  tel que

$$v_t = \alpha v_{t-1} + e_t,$$

avec  $e_t \sim i.i.d. N(0,1)$ . La question d'intérêt est de savoir si le vecteur de coefficient  $\tau$  est stable ou non et dans le dernier cas d'estimer le nombre de changements présents.

### 3.1 Détails des simulations

Dans un premier temps, nous nous intéressons à la performance des deux critères d'information lorsque aucun changement structurel n'est présent. Sans perte de généralité, le processus générateur de données est simplement un  $AR(1)$  de la forme

$$y_t = \alpha y_{t-1} + e_t,$$

avec  $e_t \sim i.i.d. N(0,1)$  et  $y_0 = 0$ . Deux cas nous intéressent. Premièrement, celui où l'on cherche à savoir s'il y a changement de moyenne, et donc  $z_t = \{1\}$ . Deuxièmement, celui où l'on cherche à savoir s'il y a changement d'ordonnée et de pente d'une fonction de tendance linéaire, et donc  $z_t = \{1, t\}$ .

Comme nous l'avons mentionné auparavant, il est possible de modéliser la dynamique d'une série de façon paramétrique (par l'introduction de variables dépendantes retardées) ou de façon non paramétrique (en laissant la corrélation dans les erreurs). Dans le premier cas, nous avons un modèle à changement structurel partiel avec  $x_t = \{y_{t-1}\}$ . Dans le deuxième cas, nous avons un modèle à changement structurel pur avec  $x_t = \{\emptyset\}$ .

Il est important de noter que la présence de corrélation dans les résidus n'empêche pas le nombre de changements sélectionné par les critères d'information d'être convergent pour la vraie valeur, pourvu que cette corrélation soit de nature stationnaire et donc, ici, que  $|\alpha| < 1$ . Dans un article récent, Nunes *et al.* (1996) ont démontré par simulations que le critère  $BIC$  tend à choisir le nombre maximal de ruptures permis lorsqu'il y a une racine unitaire ( $\alpha = 1$ ). Notre analyse s'efforcera de voir si un tel phénomène se produit encore lorsque  $|\alpha| < 1$  et comment l'importance de la corrélation affecte la performance des critères d'information.

Les autres spécifications pour les simulations sont les suivantes. La taille de l'échantillon  $T = 100$ , la longueur minimale d'un segment est  $h = 5$ , le paramètre autorégressif  $\alpha$  prend les valeurs  $\alpha = 0, 0,3, 0,5, 0,8$  et  $1$ . Le nombre maximal de ruptures varie de  $1$  à  $5$ . Finalement, les résultats sont basés sur  $1\,000$  répétitions dans chaque cas.

Dans un deuxième temps, nous répétons les mêmes exercices, mais cette fois-ci avec un changement structurel au milieu de l'échantillon, soit  $T_B = 50$ . Pour le cas d'un changement de moyenne, les données sont générées par le processus

$$\begin{aligned} y_t &= \mu \mathbf{1}(t \geq T_B) + v_t, \\ v_t &= \alpha v_{t-1} + e_t, \end{aligned} \tag{6}$$

avec  $v_0 = 0$ ,  $e_t \sim i.i.d. N(0,1)$ ,  $\mathbf{1}(\cdot)$  la fonction indicatrice. Les valeurs de  $\mu$ , le changement de moyenne, sont  $\mu = 0,5, 1$  et  $5$ . Pour le cas d'un changement de pente, les données sont générées par le processus

$$y_t = \beta (t - T_B) 1(t \geq T_B) + v_t,$$

où  $v_t$  est défini comme dans (6) et  $\beta$ , le changement de pente, prend les valeurs 0,1, 0,5 et 1.

### 3.2 Analyse des résultats

Nous commençons d'abord par décrire les résultats lorsqu'il n'y a aucun point de rupture dans le processus générateur de données. Le tableau 1 présente les résultats pour l'analyse de changement de moyenne,  $z_t = \{1\}$ , lorsque la corrélation est présente dans les résidus,  $x_t = \{\emptyset\}$ . Lorsqu'il n'y a pas de corrélation,  $\alpha = 0$ , le comportement des deux critères est satisfaisant, *BIC* choisit  $m = 0$  environ 95 % des fois et *LWZ* près de 100 %. Les choses changent lorsque le degré de corrélation s'accroît. Prenons les cas où le nombre maximal de ruptures permis est de 5. Lorsque  $\alpha = 0,5$ , *BIC* choisit  $m$  de 0 à 5 avec pratiquement la même probabilité. Lorsque  $\alpha$  s'accroît davantage, un plus grand nombre de ruptures tend à être choisi. Par exemple, avec  $\alpha = 0,7$ , plus de 50 % des fois *BIC* choisit  $m = 5$  et, si  $\alpha = 0,9$ , ce chiffre monte à 85 %. Le critère *LWZ* démontre des propriétés quelque peu similaires, mais le biais vers une surestimation du nombre de ruptures est beaucoup moins sévère. Avec  $\alpha = 0,5$ ,  $m = 0$  est choisi 76 % des fois. C'est lorsque  $\alpha$  est supérieur à 0,7 que sa performance se détériore, mais de façon moindre que pour *BIC*. Le tableau 2 montre que des résultats très semblables se produisent lorsque l'analyse porte sur la recherche de changement d'ordonnée et de pente d'une fonction de tendance,  $z_t = \{1, t\}$ .

De ces premiers résultats, on peut retenir que les deux critères *BIC* et *LWZ* conduisent aux choix satisfaisants du nombre de ruptures lorsqu'il n'y a pas de corrélation dans les résidus. Cependant, lorsque le degré de corrélation s'accroît, les critères tendent à choisir un nombre élevé de ruptures même si aucune n'est présente. Plus la corrélation s'accroît, plus les critères tendent à choisir le nombre maximal de ruptures permis. Ce biais est moindre avec *LWZ* qu'avec *BIC*, mais reste tout de même important. Ces résultats démontrent que les conclusions de Nunes *et al.* (1996) ne dépendent pas du fait que les erreurs aient une racine unitaire ; même un faible degré de corrélation mène à une surestimation.

TABLEAU 1

SÉLECTION DU NOMBRE DE RUPTURES PAR LES CRITÈRES *BIC* ET *LWZ*LORSQUE  $m_0 = 0$ ,  $z_i = \{1\}$ ,  $x_i = \{\emptyset\}$ .DONNÉES GÉNÉRÉES PAR  $y_i = \alpha * y_{i-1} + e_i$ ,  $y_0 = 0$ , $e_i \sim i.i.d. N(0,1)$ ,  $T = 100$ ,  $h = 5$ , 1 000 RÉPÉTITIONS.

(i) <i>BIC</i>	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,7	Alpha = 0,8	Alpha = 0,9	Alpha = 1,0
	1	0	96,3	46,9	20,4	7,7	2,5	0,1
		1	3,7	53,1	79,6	92,3	97,5	99,9
	2	0	92,5	24,5	2,7	0,3	0,1	0,0
		1	5,0	19,2	11,0	6,0	1,8	0,2
		2	2,5	56,3	86,3	93,7	98,1	99,8
	3	0	93,9	19,6	2,3	0,3	0,0	0,0
		1	4,8	14,2	4,9	1,4	0,2	0,0
		2	1,2	28,4	18,7	9,4	5,2	1,1
		3	0,1	37,8	74,1	88,9	94,6	98,9
	4	0	93,4	18,8	2,0	0,1	0,0	0,0
		1	4,7	12,0	3,3	0,8	0,0	0,0
		2	1,7	21,1	8,6	3,0	0,8	0,2
		3	0,2	20,7	20,3	12,0	5,9	2,6
		4	0,0	27,4	65,8	84,1	93,3	97,2
	5	0	94,6	18,2	1,8	0,1	0,0	0,0
		1	4,4	10,8	2,4	0,7	0,0	0,0
		2	0,9	21,5	6,3	2,4	0,5	0,1
		3	0,1	16,3	13,5	7,0	1,5	0,9
		4	0,0	16,9	23,0	17,0	12,6	3,2
		5	0,0	16,3	53,0	72,8	85,4	95,8

TABLEAU 1 (suite)

(ii) <i>LWZ</i>	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,7	Alpha = 0,8	Alpha = 0,9	Alpha = 1,0
	1	0	100,0	81,7	53,8	35,2	16,5	1,6
		1	0,0	18,3	46,2	64,4	83,5	98,4
	2	0	99,9	81,2	37,8	13,4	3,2	0,0
		1	0,1	11,6	24,0	22,2	14,2	3,3
		2	0,0	7,2	38,2	64,4	82,6	96,7
	3	0	99,8	78,2	34,8	12,4	2,3	0,3
		1	0,2	13,3	21,6	15,7	5,3	0,8
		2	0,0	7,1	23,2	29,3	24,7	11,1
		3	0,0	1,4	20,4	42,6	67,7	87,8
	4	0	99,7	77,9	35,6	9,9	2,2	0,0
		1	0,3	13,2	19,2	15,1	6,9	1,0
		2	0,0	6,5	22,4	24,4	14,6	4,8
		3	0,0	1,7	13,5	24,4	27,4	16,5
		4	0,0	0,7	9,3	26,2	48,9	77,7
	5	0	99,9	76,5	32,3	11,1	2,1	0,1
		1	0,1	15,2	19,0	13,2	4,5	0,9
		2	0,0	7,1	22,4	23,9	14,0	5,8
		3	0,0	1,0	15,0	21,5	21,6	11,8
		4	0,0	0,2	8,4	15,2	27,3	23,0
		5	0,0	0,0	2,9	15,1	30,5	58,4

TABLEAU 2

SÉLECTION DU NOMBRE DE RUPTURES PAR LES CRITÈRES *BIC* ET *LWZ*LORSQUE  $m_0 = 0$ ,  $z_t = \{1, t\}$ ,  $x_t = \{\emptyset\}$ .DONNÉES GÉNÉRÉES PAR  $y_t = \alpha * y_{t-1} + e_t$ ,  $y_0 = 0$ , $e_t \sim i.i.d. N(0, 1)$ ,  $T = 100$ ,  $h = 5$ , 1 000 RÉPÉTITIONS.

(i) <i>BIC</i>	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,7	Alpha = 0,8	Alpha = 0,9	Alpha = 1,0
	1	0	96,5	35,0	9,1	2,8	0,3	0,1
		1	3,5	65,0	90,9	97,2	99,7	99,9
	2	0	96,3	26,6	3,2	0,1	0,0	0,1
		1	3,4	24,6	9,8	3,2	0,6	0,6
		2	0,3	48,8	87,0	96,7	99,4	99,3
	3	0	96,6	22,5	1,6	0,1	0,0	0,0
		1	2,8	19,9	4,5	0,8	0,1	0,1
		2	0,6	25,1	16,2	7,6	3,0	0,9
		3	0,0	32,5	77,7	91,5	96,9	99,0
	4	0	95,3	21,4	1,5	0,1	0,0	0,0
		1	4,2	16,5	4,7	1,2	0,3	0,0
		2	0,5	24,0	9,0	5,4	1,2	0,2
		3	0,0	18,3	19,0	10,8	6,3	2,6
		4	0,0	19,8	65,8	82,5	92,2	97,2
	5	0	94,9	22,3	2,3	0,1	0,0	0,0
		1	4,8	20,1	5,4	1,0	0,3	0,0
		2	0,3	20,0	9,1	4,3	1,0	0,3
		3	0,0	15,1	16,3	7,9	3,2	1,1
		4	0,0	12,0	19,8	15,1	11,3	8,1
		5	0,0	10,5	47,1	71,6	84,2	90,5

TABLEAU 2 (suite)

(ii) <i>LWZ</i>	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,7	Alpha = 0,8	Alpha = 0,9	Alpha = 1,0
	1	0	100,0	85,0	50,4	25,8	8,3	1,6
		1	0,0	15,0	49,6	74,2	91,7	98,4
	2	0	100,0	85,3	41,1	16,5	4,0	0,5
		1	0,0	12,8	32,2	31,8	16,0	8,9
		2	0,0	1,9	26,7	51,7	80,0	90,6
	3	0	100,0	84,2	42,5	15,6	2,4	0,7
		1	0,0	13,6	30,6	24,6	12,8	7,4
		2	0,0	2,2	18,6	32,3	29,4	22,5
		3	0,0	0,0	8,3	27,5	55,4	69,4
	4	0	100,0	83,7	41,5	16,4	3,3	0,4
		1	0,0	13,7	31,6	26,7	12,6	6,8
		2	0,0	2,6	17,9	26,9	26,2	18,1
		3	0,0	0,0	6,9	18,8	29,7	30,7
		4	0,0	0,0	2,1	11,2	28,2	40,7
	5	0	100,0	84,4	40,6	14,9	2,4	0,2
		1	0,0	13,5	31,5	24,6	12,4	5,4
		2	0,0	1,9	19,3	28,9	27,2	16,6
		3	0,0	0,2	6,5	20,4	28,9	31,6
		4	0,0	0,0	1,8	8,3	17,6	25,5
		5	0,0	0,0	0,3	2,9	11,5	20,7

Les tableaux 3 et 4 présentent les résultats correspondants lorsque la corrélation est modélisée de façon paramétrique par l'introduction d'une variable dépendante retardée dont le coefficient n'est pas sujet à changement, c'est-à-dire avec  $x_t = \{y_{t-1}\}$ . Ici, quel que soit le degré de corrélation, même avec une racine unitaire  $\alpha = 1$ , le critère *LWZ* choisit la valeur correcte  $m = 0$  pratiquement 100 % des fois. Le critère *BIC* donne de bons résultats lorsque le degré de corrélation est moyen, mais ne choisit  $m = 0$  que 50 % des fois lorsque  $\alpha = 1$ . Ce phénomène peut être expliqué de la façon suivante : il est bien connu que dans un modèle autorégressif d'ordre 1, le coefficient autorégressif est biaisé vers le bas. Ce biais explique donc qu'il reste une certaine corrélation dans les résidus. Or, les résultats des tableaux 1 et 2 ont démontré une plus grande sensibilité à surestimer  $m$  lorsqu'il y a faible corrélation pour *BIC* que pour *LWZ*. Ces résultats tendent donc à démontrer une certaine supériorité du critère *LWZ* par rapport à *BIC* lorsqu'il y a absence de changement structurel.

TABLEAU 3

SÉLECTION DU NOMBRE DE RUPTURES PAR LES CRITÈRES *BIC* ET *LWZ*LORSQUE  $m_0 = 0$ ,  $z_t = \{1\}$ ,  $x_t = \{y_{t-1}\}$ .DONNÉES GÉNÉRÉES PAR  $y_t = \alpha * y_{t-1} + e_t$ ,  $y_0 = 0$ ,  
 $e_t \sim i.i.d. N(0,1)$ ,  $T = 100$ ,  $h = 5$ , 200 RÉPÉTITIONS.

(i) <i>BIC</i>	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 1,0
1		0	95,5	93,0	59,5
		1	4,5	7,0	40,5
2		0	92,5	89,5	49,5
		1	5,5	6,0	15,0
		2	2,0	4,5	35,5
3		0	92,0	90,0	42,5
		1	5,5	5,5	9,5
		2	2,0	1,5	15,5
		3	0,5	3,0	32,5
4		0	93,0	93,5	49,5
		1	3,5	2,0	2,0
		2	3,0	1,5	12,0
		3	0,5	2,5	15,5
		4	0,0	0,5	21,0
5		0	93,5	93,0	55,5
		1	4,5	0,5	1,0
		2	1,0	3,0	7,0
		3	1,0	2,0	5,5
		4	0,0	1,5	15,0
		5	0,0	0,0	16,0



TABLEAU 3 (suite)

(ii) <i>LWZ</i>	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 1,0
	1	0	100,0	100,0	97,5
		1	0,0	0,0	2,5
	2	0	99,5	100,0	99,0
		1	0,5	0,0	1,0
		2	0,0	0,0	0,0
	3	0	100,0	100,0	98,5
		1	0,0	0,0	1,0
		2	0,0	0,0	0,5
		3	0,0	0,0	0,0
	4	0	100,0	100,0	100,0
		1	0,0	0,0	0,0
		2	0,0	0,0	0,0
		3	0,0	0,0	0,0
		4	0,0	0,0	0,0
	5	0	100,0	100,0	99,5
		1	0,0	0,0	0,5
		2	0,0	0,0	0,0
		3	0,0	0,0	0,0
		4	0,0	0,0	0,0
		5	0,0	0,0	0,0

TABLEAU 4

SÉLECTION DU NOMBRE DE RUPTURES PAR LES CRITÈRES *BIC* ET *LWZ*LORSQUE  $m_0 = 0$ ,  $z_t = \{1, t\}$ ,  $x_t = \{y_{t-1}\}$ .DONNÉES GÉNÉRÉES PAR  $y_t = \alpha * y_{t-1} + e_t$ ,  $y_0 = 0$ ,  
 $e_t \sim i.i.d. N(0,1)$ ,  $T = 100$ ,  $h = 5$ , 200 RÉPÉTITIONS.

(i) <i>BIC</i>	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 1,0
1		0	93,0	92,5	56,5
		1	7,0	7,5	43,5
2		0	95,5	91,5	46,0
		1	3,5	6,0	14,5
		2	1,0	2,5	39,5
3		0	92,0	92,0	41,5
		1	6,5	3,0	6,5
		2	0,5	4,0	17,0
		3	1,0	1,0	32,0
4		0	98,0	92,0	50,0
		1	1,0	4,0	3,5
		2	1,0	0,5	8,5
		3	0,0	3,0	14,5
		4	0,0	0,5	23,5
5		0	95,5	94,5	56,0
		1	3,0	3,5	2,5
		2	1,0	0,5	3,5
		3	0,5	0,0	5,5
		4	0,0	1,0	12,0
		5	0,0	0,5	20,5

TABLEAU 4 (suite)

(ii) <i>LWZ</i>	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 1,0
	1	0	100,0	100,0	99,5
		1	0,0	0,0	0,5
	2	0	100,0	100,0	98,0
		1	0,0	0,0	2,0
		2	0,0	0,0	0,0
	3	0	100,0	100,0	100,0
		1	0,0	0,0	0,0
		2	0,0	0,0	0,0
		3	0,0	0,0	0,0
	4	0	100,0	100,0	99,5
		1	0,0	0,0	0,5
		2	0,0	0,0	0,0
		3	0,0	0,0	0,0
		4	0,0	0,0	0,0
	5	0	100,0	100,0	100,0
		1	0,0	0,0	0,0
		2	0,0	0,0	0,0
		3	0,0	0,0	0,0
		4	0,0	0,0	0,0
		5	0,0	0,0	0,0

Nous nous penchons maintenant sur les résultats des simulations lorsqu'il y a effectivement une rupture dans le processus générateur des données. Les résultats sont d'abord présentés dans les tableaux 5a et 5b pour le cas d'un changement de moyenne et dans les tableaux 6a et 6b pour le cas d'un changement de pente.

TABLEAU 5A

SÉLECTION DU NOMBRE DE RUPTURES PAR LE CRITÈRE *BIC*LORSQUE  $m_0 = 1$ ,  $z_t = \{1\}$ ,  $x_t = \{\emptyset\}$ .

DONNÉES GÉNÉRÉES PAR  $y_t = \mu * 1(t \geq T_B) + v_t$ ,  $v_t = \alpha * v_{t-1} + e_t$ ,  $v_0 = 0$ ,  
 $e_t \sim i.i.d. N(0,1)$ ,  $T = 100$ ,  $T_B = 50$ ,  $h = 7$ , 1 000 RÉPÉTITIONS.

(i) $\mu = 0,5$	$m$	$m$	Alpha	Alpha	Alpha	Alpha
	maximal		= 0,0	= 0,5	= 0,8	= 1,0
1		0	48,6	32,0	8,5	0,0
		1	51,4	68,0	91,5	100,0
2		0	50,4	16,3	0,5	0,0
		1	45,3	30,6	6,9	0,4
		2	4,3	53,1	92,6	99,6
3		0	48,2	14,9	0,3	0,0
		1	47,8	22,0	1,4	0,0
		2	3,9	30,5	12,3	1,7
		3	0,1	32,6	86,0	98,3
4		0	49,1	14,1	0,1	0,0
		1	46,6	20,5	1,6	0,0
		2	3,8	24,0	5,5	0,3
		3	0,5	21,0	17,8	4,9
		4	0,0	20,4	75,0	94,8
5		0	55,1	14,1	0,6	0,0
		1	42,0	22,5	0,6	0,1
		2	2,8	26,1	3,8	0,1
		3	0,1	15,9	9,1	1,4
		4	0,0	12,7	24,0	9,0
		5	0,0	8,7	61,9	89,4
(ii) $\mu = 1,0$	$m$	$m$	Alpha	Alpha	Alpha	Alpha
	maximal		= 0,0	= 0,5	= 0,8	= 1,0
1		0	1,5	6,9	7,2	0,2
		1	98,5	93,1	92,8	99,8
2		0	1,8	3,5	0,5	0,0
		1	91,2	41,0	7,0	0,8
		2	7,0	55,5	92,5	99,2
3		0	1,5	4,0	0,3	0,0
		1	89,8	30,5	2,0	0,0
		2	8,0	29,1	12,8	0,7
		3	0,7	36,4	84,9	99,3

TABLEAU 5A (suite)

(ii) $\mu = 1,0$	$m$ maximal	$m$	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	4	0	1,7	2,5	0,1	0,0
		1	92,1	29,1	0,8	0,1
		2	5,9	24,7	3,7	0,4
		3	0,3	21,7	17,6	4,7
		4	0,0	22,0	77,8	94,8
	5	0	0,9	2,8	0,3	0,0
		1	91,3	27,2	0,5	0,1
		2	7,1	22,2	3,9	0,3
		3	0,5	23,1	10,5	1,9
		4	0,0	15,4	23,0	6,4
		5	0,2	9,3	61,8	91,3
(iii) $\mu = 5,0$	$m$ maximal	$m$	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,0	0,0	0,2
		1	100,0	100,0	100,0	99,8
	2	0	0,0	0,0	0,0	0,0
		1	93,4	39,5	5,0	0,8
		2	6,6	60,5	95,0	99,2
	3	0	0,0	0,0	0,0	0,0
		1	93,8	29,5	0,7	0,0
		2	5,5	27,4	8,4	1,6
		3	0,7	43,1	90,9	98,4
	4	0	0,0	0,0	0,0	0,0
		1	91,8	25,1	0,3	0,0
		2	6,9	26,0	3,4	0,1
		3	1,3	26,3	16,2	4,2
		4	0,0	22,6	80,1	95,7
	5	0	0,0	0,0	0,0	0,0
		1	93,5	24,1	0,5	0,1
		2	5,9	23,7	3,5	0,0
		3	0,5	23,8	7,9	1,5
		4	0,1	17,1	21,3	7,4
		5	0,0	11,3	66,8	91,0

TABLEAU 5B

SÉLECTION DU NOMBRE DE RUPTURES PAR LE CRITÈRE *LWZ*

LORSQUE  $m_0 = 1$ ,  $z_t = \{1\}$ ,  $x_t = \{\emptyset\}$ .

DONNÉES GÉNÉRÉES PAR  $y_t = \mu * 1(t \geq T_B) + v_t$ ,  $v_t = \alpha * v_{t-1} + e_t$ ,  $v_0 = 0$ ,  
 $e_t \sim i.i.d. N(0,1)$ ,  $T = 100$ ,  $T_B = 50$ ,  $h = 7$ , 1 000 RÉPÉTITIONS.

(i) $\mu = 0.5$	$m$	$m$	Alpha	Alpha	Alpha	Alpha
	maximal		= 0,0	= 0,5	= 0,8	= 1,0
	1	0	88,5	64,9	31,5	1,5
		1	11,5	35,1	68,5	98,5
	2	0	89,5	59,9	12,9	0,0
		1	10,5	29,7	24,8	4,5
		2	0,0	10,4	62,3	95,5
	3	0	87,7	60,2	11,5	0,0
		1	12,3	29,5	16,7	1,7
		2	0,0	8,3	29,7	11,2
		3	0,0	2,0	42,1	87,1
	4	0	89,9	63,2	11,3	0,0
		1	10,1	26,2	15,2	1,2
		2	0,0	8,9	25,5	6,3
		3	0,0	1,3	25,0	20,6
		4	0,0	0,4	23,0	71,9
	5	0	90,7	59,9	9,8	0,1
		1	9,3	29,4	13,2	1,4
		2	0,0	8,5	24,1	4,9
		3	0,0	1,9	22,5	15,3
		4	0,0	0,3	18,6	25,5
		5	0,0	0,0	11,8	52,8
(ii) $\mu = 1,0$	$m$ maximal	$m$	Alpha	Alpha	Alpha	Alpha
			= 0,0	= 0,5	= 0,8	= 1,0
	1	0	16,2	27,2	26,0	1,4
		1	83,8	72,8	74,0	98,6
	2	0	16,9	26,1	11,2	0,0
		1	82,6	60,6	24,9	4,6
		2	0,5	13,3	63,9	95,4
	3	0	16,4	23,5	10,9	0,3
		1	83,2	62,1	17,7	2,7
		2	0,4	11,2	30,7	9,9
		3	0,0	3,2	40,7	87,1

TABLEAU 5B (suite)

(ii) $\mu = 1,0$	$m$ maximal	$m$	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	4	0	15,7	24,3	9,0	0,0
		1	84,2	60,3	15,6	1,0
		2	0,1	12,8	24,3	5,7
		3	0,0	2,1	24,5	21,4
		4	0,0	0,5	26,6	71,9
	5	0	13,9	23,9	8,0	0,0
		1	85,6	60,3	17,8	1,2
		2	0,5	11,7	23,3	4,8
		3	0,0	3,6	23,0	14,6
		4	0,0	0,5	16,4	28,4
		5	0,0	0,0	11,5	51,0
(iii) $\mu = 5,0$	$m$ maximal	$m$	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,0	0,0	1,9
		1	100,0	100,0	100,0	98,1
	2	0	0,0	0,0	0,0	0,1
		1	99,7	80,1	28,4	4,9
		2	0,3	19,9	71,6	95,0
	3	0	0,0	0,0	0,0	0,0
		1	100,0	79,8	16,5	1,0
		2	0,0	15,9	31,0	11,8
		3	0,0	4,3	52,5	87,2
	4	0	0,0	0,0	0,0	0,1
		1	99,8	78,0	13,7	0,7
		2	0,2	16,1	24,1	3,4
		3	0,0	4,8	30,5	19,6
		4	0,0	1,1	31,7	76,2
	5	0	0,0	0,0	0,0	0,0
		1	99,9	78,7	17,5	1,4
		2	0,1	16,9	22,1	5,2
		3	0,0	3,6	25,0	12,2
		4	0,0	0,5	18,9	26,5
		5	0,0	0,3	16,5	54,7

TABLEAU 6A

SÉLECTION DU NOMBRE DE RUPTURES PAR LE CRITÈRE *BIC*

LORSQUE  $m_0 = 1, z_t = \{1, t\}, x_t = \{\emptyset\}$ .

DONNÉES GÉNÉRÉES PAR  $y_t = b\hat{e}ta * (t - T_B)1(t \geq T_B) + v_t, v_t = alpha * v_{t-1} + e_t,$   
 $v_0 = 0, e_t \sim i.i.d. N(0,1), T = 100, T_B = 50, h = 7, 1\ 000$  RÉPÉTITIONS.

(i) bêta = 0,1	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,6	0,6	0,2
		1	100,0	99,4	99,4	99,8
	2	0	0,0	0,2	0,1	0,0
		1	94,9	31,0	3,4	0,2
		2	5,1	68,8	96,5	99,8
	3	0	0,0	0,2	0,0	0,0
		1	94,7	28,7	1,3	0,1
		2	5,2	32,5	6,7	1,4
		3	0,1	38,6	92,0	98,5
	4	0	0,0	0,1	0,0	0,0
		1	93,4	24,7	0,7	0,1
		2	6,1	29,4	4,5	1,2
		3	0,3	23,2	14,3	4,7
		4	0,2	22,6	80,5	94,0
	5	0	0,0	0,1	0,0	0,0
		1	95,0	27,1	0,8	0,0
		2	4,3	30,5	3,8	0,3
		3	0,7	19,4	9,3	2,7
		4	0,0	13,0	20,2	9,7
		5	0,0	9,9	65,9	87,3
(ii) bêta = 0,5	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,0	0,0	0,0
		1	100,0	100,0	100,0	100,0
	2	0	0,0	0,0	0,0	0,0
		1	93,8	20,7	1,6	0,0
		2	6,2	79,3	98,4	100,0
	3	0	0,0	0,0	0,0	0,0
		1	93,3	18,4	0,8	0,1
		2	6,1	31,6	5,6	1,0
		3	0,6	50,0	93,6	98,9



TABLEAU 6A (suite)

(ii) bêta = 0,5	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	4	0	0,0	0,0	0,0	0,0
		1	93,0	20,8	0,3	0,0
		2	5,8	26,5	4,5	0,1
		3	1,2	25,4	11,4	4,0
		4	0,0	27,3	83,8	95,9
	5	0	0,0	0,0	0,0	0,0
		1	91,9	17,2	0,3	0,0
		2	7,7	25,3	3,0	0,4
		3	0,4	23,9	8,4	1,4
		4	0,0	19,6	18,9	10,0
		5	0,0	14,0	69,4	88,2
(iii) bêta = 1,0	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,0	0,0	0,0
		1	100,0	100,0	100,0	100,0
	2	0	0,0	0,0	0,0	0,0
		1	93,7	23,8	1,8	0,3
		2	6,3	76,2	98,2	99,7
	3	0	0,0	0,0	0,0	0,0
		1	96,5	19,5	0,4	0,1
		2	3,3	30,8	4,1	0,8
		3	0,2	49,7	95,5	99,1
	4	0	0,0	0,0	0,0	0,0
		1	94,6	18,6	0,6	0,0
		2	5,1	28,2	3,4	0,3
		3	0,3	24,8	10,8	3,0
		4	0,0	28,4	85,2	96,7
	5	0	0,0	0,0	0,0	0,0
		1	94,6	20,1	0,4	0,0
		2	5,3	27,0	2,4	0,1
		3	0,1	20,1	7,2	2,3
		4	0,0	16,0	16,9	9,3
		5	0,0	16,8	73,1	88,3

TABLEAU 6B

SÉLECTION DU NOMBRE DE RUPTURES PAR LE CRITÈRE *LWZ*LORSQUE  $m_0 = 1$ ,  $z_t = \{1, t\}$ ,  $x_t = \{\emptyset\}$ .DONNÉES GÉNÉRÉES PAR  $y_t = \text{bêta} * (t - T_B)1(t \geq T_B) + v_t$ ,  $v_t = \text{alpha} * v_{t-1} + e_t$ ,  
 $v_0 = 0$ ,  $e_t \sim i.i.d. N(0,1)$ ,  $T = 100$ ,  $T_B = 50$ ,  $h = 7$ , 1 000 RÉPÉTITIONS.

(i) bêta = 0,1	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	1,2	6,2	9,7	1,9
		1	98,8	93,8	90,3	98,1
	2	0	1,1	6,1	5,1	0,6
		1	98,9	83,0	37,7	8,7
		2	0,0	10,9	57,2	90,7
	3	0	0,7	6,1	6,1	0,5
		1	99,3	82,5	29,6	7,0
		2	0,0	10,4	33,9	22,8
		3	0,0	1,0	30,4	69,7
	4	0	1,2	6,3	5,2	0,6
		1	98,7	85,0	30,3	5,4
		2	0,1	8,3	32,1	20,1
		3	0,0	0,3	20,3	31,7
		4	0,0	0,1	12,1	42,2
	5	0	1,7	5,6	4,1	0,9
		1	98,3	84,1	31,7	5,5
		2	0,0	9,5	32,1	18,4
		3	0,0	0,8	21,3	28,6
		4	0,0	0,0	8,2	28,4
		5	0,0	0,0	2,6	18,2
(ii) bêta = 0,5	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,0	0,0	0,1
		1	100,0	100,0	100,0	99,9
	2	0	0,0	0,0	0,0	0,0
		1	99,9	81,9	25,4	5,6
		2	0,1	18,1	74,6	94,4
	3	0	0,0	0,0	0,0	0,1
		1	99,9	81,1	22,9	4,1
		2	0,1	17,4	36,8	17,6
		3	0,0	1,5	40,3	78,2

TABLEAU 6B (suite)

(ii) bêta = 0,5	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	4	0	0,0	0,0	0,0	0,0
		1	100,0	82,2	22,7	3,0
		2	0,0	15,8	34,4	18,4
		3	0,0	1,7	26,8	32,9
		4	0,0	0,3	16,1	45,7
	5	0	0,0	0,0	0,0	0,0
		1	100,0	81,1	20,3	3,6
		2	0,0	16,3	36,6	14,7
		3	0,0	2,5	26,0	29,3
		4	0,0	0,1	13,5	30,5
		5	0,0	0,0	3,6	21,9
(iii) bêta = 1,0	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,0	0,0	0,0
		1	100,0	100,0	100,0	100,0
	2	0	0,0	0,0	0,0	0,0
		1	100,0	83,8	25,6	3,9
		2	0,0	16,2	74,4	96,1
	3	0	0,0	0,0	0,0	0,0
		1	100,0	83,5	20,9	3,4
		2	0,0	14,4	35,5	17,3
		3	0,0	2,1	43,6	79,3
	4	0	0,0	0,0	0,0	0,0
		1	100,0	82,4	19,3	3,5
		2	0,0	15,4	33,5	12,9
		3	0,0	2,2	27,7	34,0
		4	0,0	0,0	19,5	49,6
	5	0	0,0	0,0	0,0	0,0
		1	100,0	84,1	20,4	2,4
		2	0,0	13,8	32,6	13,7
		3	0,0	1,9	26,8	26,5
		4	0,0	0,2	15,7	29,6
		5	0,0	0,0	4,5	27,8

Prenons tout d'abord le cas où il n'y a pas de corrélation dans le processus générateur de données,  $\alpha = 0$  et  $\mu = 0,5$ . Les deux critères ont de la difficulté à détecter ce changement, et cette difficulté est d'autant plus sévère pour le critère *LWZ*. En effet, pour ce dernier,  $m = 0$  est choisi plus de 90 % des fois tandis que *BIC* choisit  $m = 0$  ou  $m = 1$  de façon presque égale. Lorsque le changement de moyenne s'accroît, les deux critères sont meilleurs et choisissent  $m = 1$  plus de 90 % des fois. S'il y a de la corrélation, on remarque toujours une forte tendance à surestimer le nombre de changements. Pour *BIC*, cette surestimation ne semble pas s'atténuer lorsque le changement de moyenne s'accroît. Pour *LWZ*, il y a cependant une nette amélioration avec  $\alpha = 0,5$  lequel correspond à un degré de corrélation modéré.

Les tableaux 6a et 6b présentent les résultats lorsqu'il y a un changement de pente dans la fonction de tendance (toujours en laissant la corrélation dans les résidus, c'est-à-dire avec  $x_t = \{\emptyset\}$ ). Les performances sont satisfaisantes lorsque  $\alpha = 0$ , surtout pour *LWZ*. Lorsque  $\alpha$  s'accroît, il y a toujours tendance à surestimer, surtout pour *BIC*. Avec  $\alpha = 0,5$ , le critère *LWZ* demeure adéquat. Cette tendance à surestimer ne diminue pas avec une augmentation du changement de pente.

Nous considérons maintenant le même cas où il y a une rupture dans le processus générateur de données, mais où, maintenant, la dynamique est modélisée de façon paramétrique, c'est-à-dire avec  $x_t = \{y_{t-1}\}$ . Les résultats sont présentés dans les tableaux 7a et 7b pour le cas d'un changement de moyenne. Lorsqu'il n'y a pas de corrélation, les résultats sont semblables à ceux des tableaux 5a et 5b. Lorsque  $\alpha = 0$ , le critère *BIC* est satisfaisant avec un changement de moyenne assez grand, tandis que *LWZ* ne l'est qu'avec un changement très grand ( $\mu = 5$ ). En présence de corrélation dans le processus générateur de données, la performance de *BIC* et *LWZ* se détériore si le changement est faible ou modéré ( $\mu = 0,5$  ou 1). Il est à noter que lorsque cette corrélation s'accroît, *LWZ* tend à choisir  $m = 0$ , c'est-à-dire qu'il devient de plus en plus conservateur.

TABLEAU 7A

SÉLECTION DU NOMBRE DE RUPTURES PAR LE CRITÈRE *BIC*

LORSQUE  $m_0 = 1, z_t = \{1\}, x_t = \{y_{t-1}\}$ .

DONNÉES GÉNÉRÉES PAR  $y_t = \mu * 1(t \geq T_B) + v_t, v_t = \alpha * v_{t-1} + e_t, v_0 = 0,$   
 $e_t \sim i.i.d. N(0,1), T = 100, T_B = 50, h = 7, 200$  RÉPÉTITIONS.

(i) $\mu = 0,5$	<i>m</i>	<i>m</i>	Alpha	Alpha	Alpha	Alpha
	maximal		= 0,0	= 0,5	= 0,8	= 1,0
	1	0	52,5	84,0	88,5	51,0
		1	47,5	16,0	11,5	49,0
	2	0	40,0	68,0	78,0	46,0
		1	47,5	22,5	13,0	29,0
		2	12,5	9,5	9,0	25,0
	3	0	40,5	70,5	72,0	35,0
		1	49,5	19,0	14,0	25,0
		2	7,0	5,5	10,0	20,0
		3	3,0	5,0	4,0	20,0
	4	0	46,0	64,5	70,0	32,0
		1	41,0	21,0	10,5	24,5
		2	7,0	6,5	9,0	21,0
		3	5,0	5,0	8,0	16,0
		4	1,0	3,0	2,5	6,5
	5	0	43,0	69,5	74,0	29,0
		1	42,5	16,5	9,0	28,5
		2	9,5	7,5	6,5	17,5
		3	3,0	4,0	3,5	10,5
		4	0,5	0,5	5,0	8,0
		5	1,5	2,0	2,0	6,5
(ii) $\mu = 1,0$		<i>m</i>	Alpha	Alpha	Alpha	Alpha
		maximal	= 0,0	= 0,5	= 0,8	= 1,0
	1	0	4,0	45,0	76,0	54,0
		1	96,0	55,0	24,0	46,0
	2	0	6,0	41,0	74,5	36,0
		1	80,5	43,5	18,0	37,0
		2	13,5	15,5	7,5	27,0
	3	0	3,5	39,5	69,5	38,5
		1	78,5	37,5	15,0	25,5
		2	12,0	14,5	10,0	19,0
		3	6,0	8,5	5,5	17,0

TABLEAU 7A (suite)

(ii) $\mu = 1,0$	$m$	$m$	Alpha	Alpha	Alpha	Alpha
	maximal		= 0,0	= 0,5	= 0,8	= 1,0
	4	0	3,5	42,0	65,0	28,5
		1	75,5	36,0	18,5	25,5
		2	17,0	13,0	11,0	18,5
		3	3,0	6,5	2,0	18,5
		4	1,0	2,5	3,5	9,0
	5	0	5,5	43,5	72,0	31,0
		1	74,0	38,0	14,0	21,5
		2	12,0	8,5	6,5	16,5
		3	6,5	6,0	4,5	11,0
		4	2,0	2,5	2,5	11,5
		5	0,0	1,5	0,5	8,5
(iii) $\mu = 5,0$	1	0	0,0	0,0	1,5	48,0
		1	100,0	100,0	98,5	52,0
	2	0	0,0	0,0	1,5	35,5
		1	91,0	78,5	57,5	27,5
		2	9,0	21,5	41,0	37,0
	3	0	0,0	0,0	2,0	26,5
		1	91,0	78,5	49,0	22,0
		2	6,0	13,5	26,0	27,5
		3	3,0	8,0	23,0	24,0
	4	0	0,0	0,0	3,0	21,0
		1	88,0	81,0	50,0	18,0
		2	8,0	13,5	20,0	20,5
		3	3,0	3,5	12,0	16,0
		4	1,0	2,0	15,0	24,5
	5	0	0,0	0,0	1,0	18,5
		1	91,5	71,5	41,0	15,5
		2	8,0	16,5	27,5	17,5
		3	0,0	5,0	10,5	15,5
		4	0,5	5,5	9,5	17,0
		5	0,0	1,5	10,5	16,0

TABLEAU 7B

SÉLECTION DU NOMBRE DE RUPTURES PAR LE CRITÈRE *LWZ*LORSQUE  $m_0 = 1$ ,  $z_t = \{1\}$ ,  $x_t = \{y_{t-1}\}$ .

DONNÉES GÉNÉRÉES PAR  $y_t = \mu * 1(t \geq T_B) + v_t$ ,  $v_t = \alpha * v_{t-1} + e_t$ ,  $v_0 = 0$ ,  
 $e_t \sim i.i.d. N(0,1)$ ,  $T = 100$ ,  $T_B = 50$ ,  $h = 7$ , 200 RÉPÉTITIONS.

(i) $\mu = 0,5$	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
1		0	94,5	98,5	99,0	92,5
		1	5,5	1,5	1,0	7,5
2		0	92,0	98,5	99,0	93,0
		1	8,0	1,5	1,0	6,0
		2	0,0	0,0	0,0	1,0
3		0	94,0	99,5	97,5	92,5
		1	6,0	0,5	2,0	6,0
		2	0,0	0,0	0,5	1,5
		3	0,0	0,0	0,0	0,0
4		0	92,0	96,5	98,0	93,5
		1	8,0	3,5	2,0	5,5
		2	0,0	0,0	0,0	1,0
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
5		0	90,0	98,0	99,0	89,5
		1	10,0	2,0	1,0	10,0
		2	0,0	0,0	0,0	0,5
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
		5	0,0	0,0	0,0	0,0
(ii) $\mu = 1,0$	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
1		0	56,5	92,5	98,5	94,5
		1	43,5	7,5	1,5	5,5
2		0	47,5	92,5	99,5	93,0
		1	52,5	7,5	0,5	7,0
		2	0,0	0,0	0,0	0,0
3		0	52,5	92,5	99,5	95,5
		1	47,0	7,5	0,5	4,0
		2	0,0	0,0	0,0	0,5
		3	0,5	0,0	0,0	0,0

TABLEAU 7B (suite)

ii) $\mu = 1,0$	$m$ maximal	$m$	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	4	0	50,5	93,5	97,0	95,0
		1	47,5	5,5	3,0	5,0
		2	2,0	1,0	0,0	0,0
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
	5	0	46,5	92,0	98,0	93,5
		1	53,0	8,0	2,0	5,0
		2	0,5	0,0	0,0	1,5
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
		5	0,0	0,0	0,0	0,0
iii) $\mu = 5,0$	$m$ maximal	$m$	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,0	39,0	89,0
		1	100,0	100,0	61,0	11,0
	2	0	0,0	0,0	41,5	85,5
		1	99,5	96,5	54,5	13,0
		2	0,5	3,5	4,0	1,5
	3	0	0,0	0,0	41,0	89,5
		1	100,0	97,5	53,0	8,5
		2	0,0	2,5	5,0	1,5
		3	0,0	0,0	1,0	0,5
	4	0	0,0	0,0	42,5	89,0
		1	100,0	97,5	53,0	7,0
		2	0,0	2,5	4,0	3,0
		3	0,0	0,0	0,5	0,0
		4	0,0	0,0	0,0	1,0
	5	0	0,0	0,0	38,0	91,5
		1	100,0	99,5	53,5	5,0
		2	0,0	0,5	8,0	2,5
		3	0,0	0,0	0,5	0,5
		4	0,0	0,0	0,0	0,5
		5	0,0	0,0	0,0	0,0



Les résultats pour le cas avec changement de pente sont présentés dans les tableaux 8a et 8b. Les critères *BIC* et *LWZ* se comportent bien si  $\alpha = 0$  (cette performance s'améliore lorsque le changement s'accroît pour *LWZ*, mais non pour *BIC*). Avec de la corrélation dans le processus générateur de données ( $\alpha = 0,5$  ou  $0,8$ ), les critères *BIC* et *LWZ* deviennent conservateurs et ont tendance à choisir  $m = 0$  beaucoup plus souvent lorsque  $\alpha$  s'accroît. Ce biais vers le bas s'atténue cependant lorsque le changement de pente s'accroît.

TABLEAU 8A

SÉLECTION DU NOMBRE DE RUPTURES PAR LE CRITÈRE *BIC*LORSQUE  $m_0 = 1$ ,  $z_t = \{1, t\}$ ,  $x_t = \{y_{t-1}\}$ .

DONNÉES GÉNÉRÉES PAR  $y_t = \text{bêta} * (t - T_B)1(t \geq T_B) + v_t$ ,  $v_t = \text{alpha} * v_{t-1} + e_t$ ,  
 $v_0 = 0$ ,  $e_t \sim i.i.d. N(0,1)$ ,  $T = 100$ ,  $T_B = 50$ ,  $h = 7$ , 200 RÉPÉTITIONS.

(i) bêta = 0,1	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	30,5	63,5	52,5
		1	100,0	69,5	36,5	47,5
	2	0	0,0	22,0	57,5	38,5
		1	92,5	60,0	29,5	32,5
		2	7,5	18,0	13,0	29,0
	3	0	0,0	25,5	53,5	28,0
		1	91,0	54,0	27,0	25,5
		2	8,0	16,0	13,0	24,0
		3	1,0	4,5	6,5	22,5
	4	0	0,0	22,5	54,5	29,5
		1	90,0	62,0	24,5	24,0
		2	8,5	11,0	10,0	20,5
		3	1,5	2,5	3,5	14,5
		4	0,0	2,0	7,5	11,5
	5	0	0,0	22,5	58,5	29,0
		1	91,5	63,5	21,5	22,0
		2	7,0	9,0	9,0	12,5
		3	1,0	3,0	6,0	15,5
		4	0,5	1,0	2,5	12,0
		5	0,0	1,0	2,5	9,0
(ii) bêta = 0,5	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,0	6,0	33,5
		1	100,0	100,0	94,0	66,5

TABLEAU 8A (suite)

2	0	0,0	0,0	8,5	26,5
	1	92,5	85,0	49,5	38,5
	2	7,5	15,0	42,0	35,0
3	0	0,0	0,0	5,5	19,0
	1	84,5	79,0	56,0	36,0
	2	13,5	16,0	20,5	24,5
	3	2,0	5,0	18,0	20,5
4	0	0,0	0,0	2,5	12,5
	1	91,5	73,5	55,0	27,5
	2	6,0	18,5	24,0	24,5
	3	1,5	6,0	12,0	14,0
	4	1,0	2,0	6,5	21,5
5	0	0,0	0,0	2,0	18,0
	1	89,0	75,5	52,0	25,5
	2	9,0	17,0	22,5	18,0
	3	2,0	4,0	11,5	14,5
	4	0,0	2,0	7,0	14,0
	5	0,0	1,5	5,0	10,0

(iii) bêta = 1,0	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
1		0	0,0	0,0	2,5	18,0
		1	100,0	100,0	97,5	82,0
2		0	0,0	0,0	1,5	9,0
		1	91,5	83,0	66,0	52,0
		2	8,5	17,0	32,5	39,0
3		0	0,0	0,0	0,5	4,0
		1	90,0	88,5	51,0	39,5
		2	8,5	9,5	25,5	22,5
		3	1,5	2,0	23,0	34,0
4		0	0,0	0,0	2,0	7,5
		1	89,5	77,0	51,5	32,5
		2	9,0	16,5	20,5	23,5
		3	6,0	4,5	13,5	18,5
		4	0,5	2,0	12,5	18,0
5		0	0,0	0,0	1,0	4,5
		1	90,5	86,5	56,5	33,5
		2	7,5	6,0	19,0	23,5
		3	1,0	5,0	16,5	19,0
		4	1,0	1,0	3,5	11,0
		5	0,0	1,5	3,5	8,5

TABLEAU 8B

SÉLECTION DU NOMBRE DE RUPTURES PAR LE CRITÈRE LWZ

LORSQUE  $m_0 = 1$ ,  $z_t = \{1, t\}$ ,  $x_t = \{y_{t-1}\}$ .

DONNÉES GÉNÉRÉES PAR  $y_t = \text{bêta} * (t - T_B)1(t \geq T_B) + v_t$ ,  $v_t = \text{alpha} * v_{t-1} + e_t$ ,  
 $v_0 = 0$ ,  $e_t \sim i.i.d. N(0,1)$ ,  $T = 100$ ,  $T_B = 50$ ,  $h = 7$ , 200 RÉPÉTITIONS.

(i) bêta = 0,1	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	7,0	89,5	100,0	98,0
		1	93,0	10,5	0,0	2,0
	2	0	8,0	91,5	97,5	97,5
		1	91,5	8,5	2,5	1,5
		2	0,5	0,0	0,0	1,0
	3	0	8,5	91,0	98,0	99,5
		1	91,5	9,0	2,0	4,5
		2	0,0	0,0	0,0	0,0
		3	0,0	0,0	0,0	0,0
	4	0	4,5	90,5	98,0	97,0
		1	95,5	9,5	2,0	3,0
		2	0,0	0,0	0,0	0,0
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
	5	0	8,0	90,0	99,0	96,0
		1	92,0	9,5	1,0	4,0
		2	0,0	0,5	0,0	0,0
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
		5	0,0	0,0	0,0	0,0
(ii) bêta = 0,5	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	3,0	84,0	96,0
		1	100,0	97,0	16,0	4,0
	2	0	0,0	4,0	82,0	95,0
		1	100,0	95,5	16,5	5,0
		2	0,0	0,5	1,5	0,0
	3	0	0,0	6,5	81,5	94,0
		1	100,0	93,5	18,5	5,0
		2	0,0	0,0	0,0	1,0
		3	0,0	0,0	0,0	0,0

TABLEAU 8B (suite)

(ii) bêta = 0,5	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	4	0	0,0	3,5	83,5	94,5
		1	100,0	95,5	16,0	5,0
		2	0,0	1,0	0,5	0,5
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
	5	0	0,0	7,5	85,5	92,5
		1	99,5	92,5	14,5	6,5
		2	0,5	0,0	0,0	1,0
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
(iii) bêta = 1,0	<i>m</i> maximal	<i>m</i>	Alpha = 0,0	Alpha = 0,5	Alpha = 0,8	Alpha = 1,0
	1	0	0,0	0,5	56,5	78,0
		1	100,0	99,5	43,5	22,0
	2	0	0,0	0,0	56,0	75,0
		1	99,5	100,0	43,0	24,0
		2	0,5	0,0	1,0	1,0
	3	0	0,0	0,5	61,0	80,5
		1	100,0	99,5	38,0	19,0
		2	0,0	0,0	1,0	0,5
		3	0,0	0,0	0,0	0,0
	4	0	0,0	1,0	68,0	72,5
		1	100,0	98,5	32,0	27,5
		2	0,0	0,5	0,0	0,0
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
	5	0	0,0	0,5	57,0	77,0
		1	99,5	99,0	43,0	23,0
		2	0,5	0,5	0,0	0,0
		3	0,0	0,0	0,0	0,0
		4	0,0	0,0	0,0	0,0
		5	0,0	0,0	0,0	0,0

De l'ensemble des résultats, on peut tirer les conclusions suivantes : en premier lieu, lorsqu'il n'y a pas de corrélation dans le processus générateur de données, les deux critères *BIC* et *LWZ* donnent des résultats satisfaisants. En second lieu, lorsqu'il y a corrélation et que celle-ci n'est pas explicitement modélisée de façon paramétrique, il y a un biais très sévère à choisir un nombre élevé de changements (surtout pour *BIC*). Lorsque l'on tient compte de cette corrélation de façon paramétrique, le critère *LWZ* performe très bien si  $m = 0$ , mais a tendance à sélectionner  $m = 0$  encore, même si un changement est présent. Pour le critère *BIC*, celui-ci tend à sélectionner un nombre de changements biaisé vers le haut lorsque  $m = 0$ , mais performe mieux lorsque  $m = 1$ . Il semble donc y avoir un arbitrage entre le caractère conservateur de *LWZ*, qui fonctionne mieux lorsque le nombre de changements est faible, et le caractère libéral de *BIC*, qui lui performe mieux lorsque  $m$  est supérieur à 0.

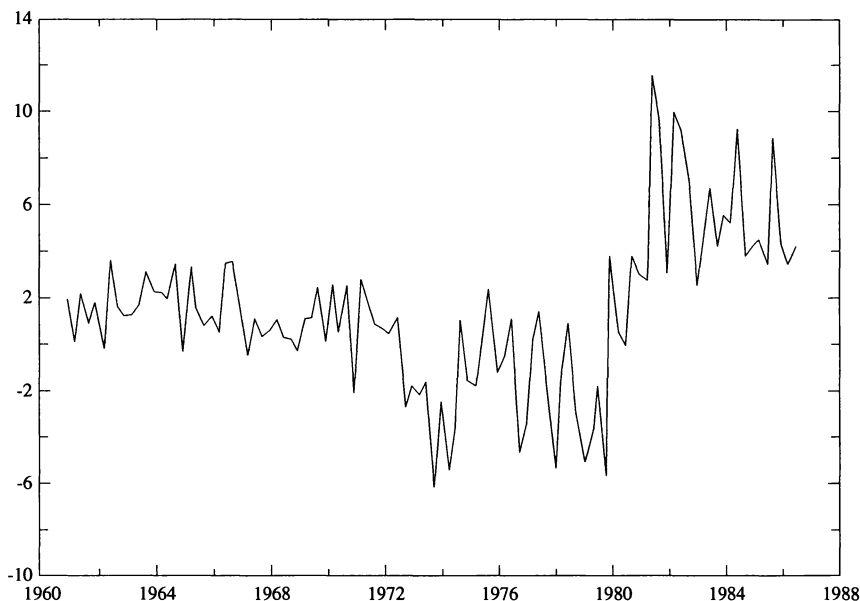
#### 4. ILLUSTRATIONS EMPIRIQUES

Dans cette section, nous présentons deux exemples empiriques pour illustrer l'importance des résultats de simulations. Le premier exemple considère le taux d'intérêt réel *ex post* des États-Unis (construit à partir du taux des bons du Trésor de trois mois dégonflé par l'indice des prix à la consommation de la banque de données Citibase). Les données sont trimestrielles de 1961:1 à 1986:3 et sont présentées à la figure 2. Cette même série a été analysée par Garcia et Perron (1996) qui ont argumenté que sa moyenne avait subi deux changements majeurs en 1973 et 1981. Cependant, une fois que ces changements sont pris en compte, la série n'exhibe pratiquement pas de corrélation temporelle, c'est-à-dire que le taux d'intérêt est essentiellement une séquence de termes indépendants.

Nous avons appliqué la procédure d'estimation discutée aux sections 1 et 2 en considérant le cas simple où  $z_t = \{1\}$ , une constante ( $q = 1$ ), et sans aucune variable  $x_t$ . Il s'agit donc d'un modèle à changement structurel pur, et la dynamique, s'il y en a, n'est pas modélisée d'une façon paramétrique. Le nombre maximal de ruptures permis est de 5 et chaque segment est contraint à avoir au moins 7 observations. Les résultats démontrent que chacun des critères d'information, *BIC* et *LWZ*, sélectionne 2 ruptures estimées à 1972:3 et 1980:3. Les différences des moyennes estimées sont significatives et démontrent une baisse de 3,16 % à la fin de 1972 et une hausse de 7,44 % à la fin de 1980. Il est important de remarquer, comme l'ont fait Garcia et Perron (1996), que la série ne présente pratiquement pas de corrélation à l'intérieur de chaque segment. En se basant sur les résultats de simulations de la section 3, on peut donc être confiant que le nombre de ruptures sélectionné est adéquat.

FIGURE 2

TAUX D'INTÉRÊT RÉEL AUX ÉTATS-UNIS; 1961:1-1986:3



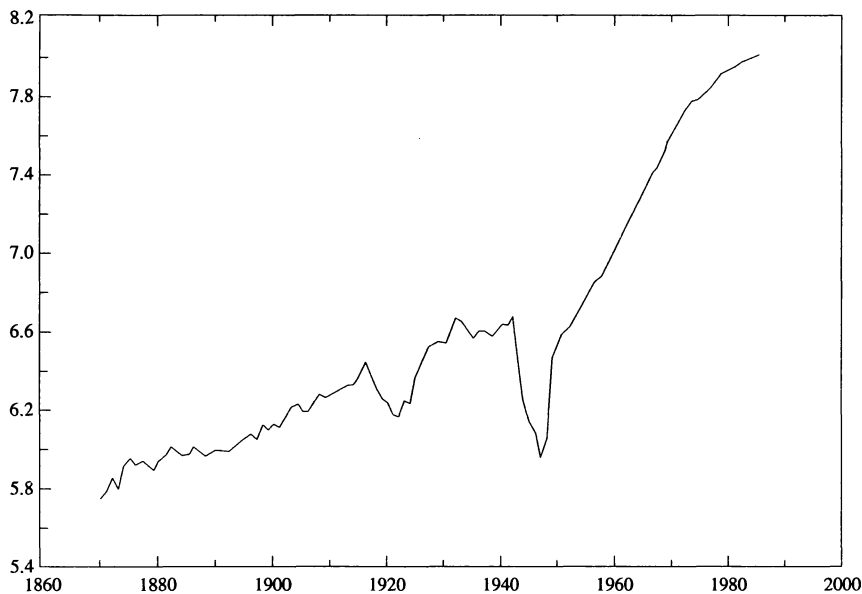
Le deuxième exemple présente un cas où la série est fortement corrélée à l'intérieur de chaque segment potentiel. Il s'agit du logarithme du PNB réel français pour la période 1870-1986. Ces données annuelles sont de Maddison (1982), complétées avec des données de l'International Financial Statistics pour l'après-guerre. L'intérêt, ici, est d'examiner les changements de la fonction de tendance, celle-ci étant modélisée à la base comme un simple polynôme d'ordre un en  $t$ .

Dans un premier temps, nous avons adopté une approche non paramétrique pour traiter la dynamique de la série. Donc, ici, les spécifications sont  $z_t = \{1, t\}$  ( $q = 2$ ) et  $x_t$  est un ensemble nul. Le nombre minimal d'observations dans chaque segment est  $h = 5$ . Les résultats démontrent que quel que soit le nombre maximal de ruptures permis, les critères  $BIC$  et  $LWZ$  indiquent tous deux de choisir ce nombre maximal comme valeur estimée. Ceci est sans aucun doute dû à la présence de forte corrélation dans les données.

Nous avons donc, dans un deuxième temps, adopté une approche paramétrique pour tenir compte de la présence de corrélation. Le vecteur  $z_t$  est toujours  $z_t = \{1, t\}$ , mais ici, le vecteur  $x_t$  comprend 6 retards de la variable, c'est-à-dire  $x_t = \{y_{t-1}, y_{t-2}, \dots, y_{t-6}\}$ . En imposant cette spécification, nous ne permettons pas aux coefficients autorégressifs de varier à travers chaque segment. Les autres spécifications sont un nombre maximal de ruptures égal à 5 et un nombre minimal d'observations dans chaque segment de  $h = 10$ . Puisqu'il s'agit d'un modèle

FIGURE 3

PNB RÉEL DE LA FRANCE; 1870-1986



à changement structurel partiel, nous devons utiliser la méthode itérative pour estimer les points de rupture correspondant aux minimums globaux des sommes des résidus au carré. Pour initialiser cette méthode récursive, nous permettons, dans un premier temps, à tous les coefficients de varier à travers chaque régime, c'est-à-dire que nous spécifions  $z_t = \{1, t, y_{t-1}, \dots, y_{t-6}\}$  et  $x_t$  l'ensemble nul (d'où le besoin d'avoir  $h$  au moins égal à 8). Nous appliquons cette procédure itérative pour  $m$  variant de 1 à 5. Il est intéressant de noter que le nombre maximal d'itérations requis est de 3.

Les résultats démontrent que les deux critères d'information *BIC* et *LWZ* s'entendent pour choisir  $m = 2$  comme estimation du nombre de ruptures. Les dates de changement sont 1943 et 1970. Les coefficients estimés démontrent une forte hausse du taux de croissance en 1943 (de 0,85 % à 3,90 % par année) et une baisse en 1970 (à 1,76 % par année). La forte hausse en 1943 est associée à une baisse marquée de la moyenne (ou ordonnée de la fonction de tendance) causée par la seconde guerre. L'estimation de la somme des coefficients autorégressifs est de 0,59, bien inférieure à la borne de racine unitaire, mais néanmoins assez élevée pour expliquer le problème encouru de la sélection du nombre maximal de ruptures lorsque la corrélation n'est pas modélisée explicitement de façon paramétrique.

## CONCLUSION

Cette étude a d'abord présenté un survol des résultats théoriques de Bai et Perron (1996) concernant l'estimation de modèles à changements structurels multiples. Ensuite, nous avons abordé en détail un algorithme efficace, basé sur une approche de programmation dynamique, qui permet d'obtenir les estimations correspondant au minimum global de la somme des résidus au carré. En présentant cet algorithme, notre but est aussi de démystifier l'idée générale que dès qu'il y a plus de deux changements, les modèles usuels deviennent vite difficiles à estimer, puisqu'il existe un nombre de  $m$ -partitions possible qui est d'ordre  $T^m$ . En fait, nous avons démontré qu'avec la capacité de calcul disponible aujourd'hui, le temps requis pour estimer ces modèles est relativement modeste même pour un grand nombre de changements et une taille d'échantillon élevée.

Notre étude a ensuite abordé le problème de la détermination du nombre de changements via certains critères d'information. Le message de nos résultats est clair, la présence de corrélation affecte grandement la précision de l'estimation. Sans corrélation, ces critères sont satisfaisants, mais se détériorent rapidement lorsque la corrélation dans les résidus s'accroît. Une modélisation paramétrique de cette corrélation est alors nécessaire. Cette dernière permet aux critères d'informations de donner des résultats relativement précis (surtout avec le critère *LWZ*), sauf lorsque la structure autorégressive a une racine près de un. Il est donc nécessaire d'appliquer ces critères avec une certaine prudence. Dans une étude ultérieure, il serait utile de comparer le mérite relatif de ces méthodes en faisant une comparaison avec la méthode suggérée par Bai et Perron (1996) basée sur une séquence de tests.

## BIBLIOGRAPHIE

- ANDREWS, D.W.K. (1991), « Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation », *Econometrica*, 59 : 817-858.
- ANDREWS, D.W.K. (1993), « Tests for Parameter Instability and Structural Change with Unknown Change Point », *Econometrica*, 61 : 821-856.
- ANDREWS, D.W.K., I. LEE, et W. PLOBERGER (1996), « Optimal Changepoint Tests for Normal Linear Regression », *Journal of Econometrics*, 70 : 9-38.
- ANDREWS, D.W.K., et W. PLOBERGER (1994), « Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative », *Econometrica*, 62 : 1383-1414.
- BAI, J. (1994a), « Least Squares Estimation of a Shift in Linear Processes », *Journal of Time Series Analysis*, 15 : 453-472.
- BAI, J. (1994b), « Estimation of Structural Change based on Wald Type Statistics », à paraître dans *Review of Economics and Statistics*.



- BAI, J., R.L. LUMSDAINE, et J.H. STOCK (1994), « Testing for and Dating Breaks in Integrated and Cointegrated Time Series », mimeo, Kennedy School of Government, Harvard University.
- BAI, J., et P. PERRON (1996), « Estimating and Testing Linear Models with Multiple Structural Changes », à paraître dans *Econometrica*.
- BAI, J., et P. PERRON (1997), « Computation and Analysis of Multiple Structural Change Models », mimeo en préparation, Centre de recherche et développement en économique, Université de Montréal.
- BARTEN, A.P., et L. SALVAS-BRONCARD (1970), « Two-Stage Least Squares Estimation with Shifts in the Structural Form », *Econometrica*, 38 : 938-941.
- BELLMAN, R., et R. ROTH (1969), « Curve Fitting by Segmented Straight Lines », *Journal of the American Statistical Association*, 64 : 1079-1084.
- BHATTACHARYA, P.K. (1994), « Some Aspects of Change-Point Analysis », dans E. CARLSTEIN, H.-G. MÜLLER et D. SIEGMUNG (eds.), *Change Point Problems*, IMS Lecture Notes - Monograph Series, vol. 23 : 28-56.
- BROWN, R.L., J. DURBIN, et J.M. EVANS (1975), « Techniques for Testing the Constancy of Regression Relationships Over Time », *Journal of the Royal Statistical Society, Series B*, 37 : 149-192.
- CHOW, G. (1960), « Tests of Equality Between Sets of Regression Coefficients in Two Linear Regressions », *Econometrica*, 28 : 591-605.
- DESHAYES, J., et D. PICARD (1986), « Off-line Statistical Analysis of Change Point Models Using Non Parametric and Likelihood Methods », dans M. BASSEVILLE et A. BENVENISTE (eds.), *Detection of Abrupt Changes in Signals and Dynamical Systems*, Springer Verlag Lecture Notes in Control and Information Sciences No. 77 : 103-168.
- FISHER, W.D. (1958), « On Grouping for Maximum Homogeneity », *Journal of the American Statistical Association*, 53 : 789-798.
- GARCIA, R., et P. PERRON (1996), « An Analysis of the Real Interest Rate under Regime Shifts », *Review of Economics and Statistics*, 78 : 111-125.
- GUTHERY, S.B. (1974), « Partition Regression », *Journal of the American Statistical Association*, 69 : 945-947.
- KRISHNAIAH, P.R., et B.Q. MIAO (1988), « Review about Estimation of Change Points », dans *Handbook of Statistics*, vol. 7, P.R. KRISHNAIAH et C.R. RAO (eds.), New York : Elsevier.
- LIU, J., S. WU, et J.V. ZIDEK (1995), « On Segmented Multivariate Regressions », mimeo, Département de Statistiques, University of British Columbia.
- MADDISON, A. (1982), *Phases of Capitalist Development*. London : Oxford University Press.
- NEWKEY, W.K., et K.D. WEST (1987), « A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix », *Econometrica*, 55 : 703-708.
- NUNES, L.C., P. NEWBOLD, et C.-M. KUAN (1996), « Spurious Number of Breaks », *Economics Letters*, 50 : 175-178.

- PERRON, P. (1989), « The Great Crash, the Oil Price Shock and the Unit Root Hypothesis », *Econometrica*, 57 : 1361-1401.
- PERRON, P. (1991), « A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series », mimeo, Département de Sciences économiques, Princeton University.
- PERRON, P. (1994), « Trend, Unit Root and Structural Change in Macroeconomic Time Series », dans *Cointegration for the Applied Economists*, B.B. RAO (ed.), MacMillan Press : 113-146.
- QUANDT, R.E. (1958), « The Estimation of Parameters of a Linear Regression System Obeying Two Separate Regimes », *Journal of the American Statistical Association*, 55 : 873-880.
- SARGAN, J.D. (1964), « Wages and Prices in the United Kingdom: A Study in Econometric Methodology », dans *Econometric Analysis for National Economic Planning*, P.E. HART, G. MILLS et J.K. WHITAKER (eds.), London : Butterworths : 25-54.
- SOWELL, F. (1996), « Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework », *Econometrica*, 64 : 1085-1107.
- VOGELSANG, T.J. (1993), « Wald-type Tests for Detecting Shifts in the Trend Function of a Dynamic Time Series », mimeo, Département de Sciences économiques, Cornell University.
- YAO, Y.-C. (1988), « Estimating the Number of Change-Points via Schwarz' Criterion », *Statistics and Probability Letters*, 6 : 181-189.
- YAO, Y.-C., et S.T. AU (1989), « Least-Squares Estimation of a Step Function », *Sankhyā*, 51, Series A : 370-381.
- YIN, Y.Q. (1988), « Detection of the Number, Locations and Magnitudes of Jumps », *Communications in Statistics-Stochastic Models*, 4 : 445-455.